

1	PRIOR APPLICATION NUMBER: US 60/219,556
2	PRIOR FILING DATE: 2000-07-20
3	PRIOR APPLICATION NUMBER: US 60/220,624
4	PRIOR FILING DATE: 2000-07-25
5	PRIOR APPLICATION NUMBER: US 60/220,664
6	PRIOR FILING DATE: 2000-07-25
7	PRIOR APPLICATION NUMBER: PCT/US00/20710
8	PRIOR FILING DATE: 2000-07-28
9	PRIOR APPLICATION NUMBER: US 60/222,695
10	PRIOR FILING DATE: 2000-08-02
11	PRIOR APPLICATION NUMBER: US 09/643,657
12	PRIOR FILING DATE: 2000-08-17
13	PRIOR APPLICATION NUMBER: PCT/US00/23522
14	PRIOR FILING DATE: 2000-08-23
15	PRIOR APPLICATION NUMBER: PCT/US00/23328
16	PRIOR FILING DATE: 2000-08-24
17	PRIOR APPLICATION NUMBER: US 60/230,978
18	PRIOR FILING DATE: 2000-09-07
19	PRIOR APPLICATION NUMBER: US 60/000,000
20	PRIOR FILING DATE: 2000-09-15
21	PRIOR APPLICATION NUMBER: US 09/664,610
22	PRIOR FILING DATE: 2000-09-18
23	PRIOR APPLICATION NUMBER: US 09/665,350
24	PRIOR FILING DATE: 2000-09-18
25	PRIOR APPLICATION NUMBER: US 60/242,922
26	PRIOR FILING DATE: 2000-10-24
27	PRIOR APPLICATION NUMBER: US 09/709,238
28	PRIOR FILING DATE: 2000-11-08
29	PRIOR APPLICATION NUMBER: PCT/US00/30952
30	PRIOR FILING DATE: 2000-11-08
31	PRIOR APPLICATION NUMBER: PCT/US00/30873
32	PRIOR FILING DATE: 2000-11-10
33	PRIOR APPLICATION NUMBER: PCT/US00/32678
34	PRIOR FILING DATE: 2000-12-01
35	PRIOR APPLICATION NUMBER: US 09/747,259
36	PRIOR FILING DATE: 2000-12-20
37	PRIOR APPLICATION NUMBER: PCT/US00/34956
38	PRIOR FILING DATE: 2000-12-20
39	PRIOR APPLICATION NUMBER: US 09/767,609
40	PRIOR FILING DATE: 2001-01-22
41	PRIOR APPLICATION NUMBER: US 09/796,498
42	PRIOR FILING DATE: 2001-02-28
43	PRIOR APPLICATION NUMBER: PCT/US01/06520
44	PRIOR FILING DATE: 2001-02-28
45	PRIOR APPLICATION NUMBER: PCT/US01/06666
46	PRIOR FILING DATE: 2001-03-01
47	PRIOR APPLICATION NUMBER: US 09/802,706
48	PRIOR FILING DATE: 2001-03-09
49	PRIOR APPLICATION NUMBER: US 09/808,689
50	PRIOR FILING DATE: 2001-03-14
51	PRIOR APPLICATION NUMBER: US 09/816,744
52	PRIOR FILING DATE: 2001-03-22
53	PRIOR APPLICATION NUMBER: US 09/828,366
54	PRIOR FILING DATE: 2001-04-05
55	PRIOR APPLICATION NUMBER: US 09/854,208
56	PRIOR FILING DATE: 2001-05-10
57	PRIOR APPLICATION NUMBER: US 09/854,280
58	PRIOR FILING DATE: 2001-05-10
59	PRIOR APPLICATION NUMBER: US 09/866,028
60	PRIOR FILING DATE: 2001-05-25
61	PRIOR APPLICATION NUMBER: US 09/866,034
62	PRIOR FILING DATE: 2001-05-25
63	PRIOR APPLICATION NUMBER: PCT/US01/17092
64	PRIOR FILING DATE: 2001-05-25
65	PRIOR APPLICATION NUMBER: US 09/870,574
66	PRIOR FILING DATE: 2001-05-30
67	PRIOR APPLICATION NUMBER: PCT/US01/17443
68	PRIOR FILING DATE: 2001-05-30
69	PRIOR APPLICATION NUMBER: PCT/US01/17800
70	PRIOR FILING DATE: 2001-06-01
71	PRIOR APPLICATION NUMBER: PCT/US01/19692
72	PRIOR FILING DATE: 2001-06-20
73	PRIOR APPLICATION NUMBER: PCT/US01/20000

```

: : PRIOR FILING DATE: 2001-06-28
: :
: : NUMBER OF SEQ ID NOS: 383
: :
: : SEQ ID NO 163
: :
: : LENGTH: 1049
: :
: : TYPE: DNA
: :
: : ORGANISM: Homosapiens
US-10-081-056-163

```

Query Match		96.3%	Score 990.4	DB 37	Length 1049
Best Local Similarity		99.4%	Prod. No. 5.8e-269		
Matches	994	Conservative	0	Mismatches	6
				Indels	0
QY	25	TCTAAACGGTGTCAAGACAAAGATGCTTCAGCTTTGGAAACTGTTCCTCTCGCGGC	84		
DB	43	tctgagggtggtgtcaadacaaaagaTgtcttcagcttggaaacttgttctctctgcggc	102		
QY	85	GTGCTCACTGGGACCTCAGAGTCTCTTCTTGCAATCTTGGCAATGACCTAAGCAATGTC	144		
DB	103	gtgctcactgggaactcagagtcctcttctgacaattctggcaatgacctaaagcaatgctc	162		
QY	145	GTGGATAAGCTTGGAACTGTCTTCAGGAGGACTTGACAGAGTTGACAACTCTTAAA	204		
DB	163	gtggataagctggaaactgttcttcagaggagcttggacagttgacaaactactctaaa	222		
QY	205	GGCATCTCTTGAAACTGAAGGTGCAGCTAGGAGTGTCTCAGAAATCAGAGTGTGCCAA	264		
DB	223	ggcatctctgaaaaactgaaggtgcagactaggagtgcttcagaaatcagtgcttgcaa	282		
QY	265	CTGGCCAGCAGAGAGGCCAGGAGCTGACAAATTCCTTGACAAATGTCATTTCCTAAGCTG	324		
DB	283	ctggccagcagagggccaggagcttgagaaatTgtgaaacaTgtcattctctaaagctg	342		
QY	325	CTTCCAACCTAACACGGACATTTTGGGTTGAAATCAGCAACTCCCTCATCTCGATGTC	384		
DB	343	cttccaacttaacacggacatttttgggttgaaaatcagaactccctcatctggatgctc	402		
QY	385	AAAGCTGAACGATCGATGATGCGCAAGGCCCTTAACCTGAGCTTCCCTGTACCGCCGAAT	444		
DB	403	aaagctgaaccgactcgatgatgacaggagccttaacctgaagcttccctgtcacgcgaat	462		
QY	445	GTCAGTGTGGCGGGCCCATCATTTGGCCAGATATCAAGCTGAAAGCCCTCTTGAGACTC	504		
DB	463	gtcactgtggcggggcccatcatttggcgagattatcaacctgaaagcctcttggaactc	522		
QY	505	CTGACCGCAGTTCACAAATGAACTGATGCCAGACACACCAGCTTGTTCCTGCTCTGGGA	564		
DB	523	ctgacgcagcttcaaatTgaactgataccccagacacacacagctgttTgcgctctggga	582		
QY	565	GAATGGCCAGTGAACCAACAGCATCTCACTTCTCTGCTGGACAAACACAGCCAATC	624		
DB	583	gaatggccagatgacccaacagcatctcaacttctctgtggacaaacacagccaactc	642		
QY	625	ATCACACAGTTCGTGAATAGCGGTATCACAGCTGAAAGACTGTATCTCCTCGCTG	684		
DB	643	atcaacaagttTcgtgaatagcgtgatacaacgctgaaagcactgtatctccctcgtg	702		
QY	685	CAGAAGGAGATATGTCCACTGTATCCGCACTCTTCATCCACTCCCTGGATGTGAATGTCTT	744		
DB	703	cagaaggagatatgtccactgataccgactctcatccactccctggatgtgaatgtcaatt	762		
QY	745	CAGCAGTGTCTGATAATCTCTAGCACAAACCCAGCTGCAACCCCTCATTTTGAAGAGGA	804		
DB	763	cagcaggtTcgtcgaataatctcagcaaaaaacccagctgcaaacctctatctgaagagga	822		
QY	805	CGAATGAGGAGGACACTGTGGTGCATGCTGTATGTTGTTCCGAGTGGCTGCCACCCCC	864		
DB	823	cgaaTgaaggaggaactcgtTggtgcatgctgattggttcccaagtggcttggccaccccc	882		
QY	865	TTATTAGCATCTCCCTCCAGGAAGTCTGCTGCCACCACTTAACACAGCGTGAAGCTGAGTC	924		
DB	883	ttaTgaactctccctccaggaagctgtctgcacacactaaacacagctTgaagcgtctgaatc	942		

```

QY 925 CCACGAGAGGACCTTCCAGATACCCCTTCTCTCAGTCAGCAACAGCAGCCTCTTACA 984
Db 943 CCACGAGAGGACCTTCCAGATACCCCTTCTCTCAGTCAGCAACAGCAGCCTCTTACA 1002
QY 985 CATGTTCTCTCTCCCTCGCCCAATAAGGCCCATTTCTTCGA 1024
Db 1003 catgtgtctctgccccctggcaataaaggcccatcttcgca 1042

RESULT 11
US-60-324-185-33384
; Sequence 33384, Application US/60324185
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE SEQUENCES
; FILE REFERENCE: GX-0019-1 P
; CURRENT APPLICATION NUMBER: US/60/324,185
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 35862
; SOFTWARE: PERL Program
; SEQ ID NO 33384
; LENGTH: 1106
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte ID No: 903509.6
US-60-124-185-33384

```

Query Match	93.5%	Score	961.4	DB	71	Length	1106
Best Local Similarity	98.9%	Pred.	No.	9.6e-361			
Matches	968	Conservative	0	Mismatches	11	Indels	0
Gaps	0						
QY	25	TCTAACGCGTCTCAAGACAAAGATGCTTCAGCTTTTGGAAACTGTTCTCTCTGTGCGGC	84				
Db	43	tctgagggtgtctcaagacaaaagatgcttcagcttctggaacttctctctctgtgcgac	102				
QY	85	GTGCTCAGTCGGACGCTCAGAGTCTCTTTGACAATTTGGCAATGACCTTAAGCAATGTC	144				
Db	103	gtgtcaactggaaacctcagagtcctctcttgacaactctggcaatctggcaatgaacaaatg	162				
QY	145	GTGGATAAGCTGGACACTGTTCTTCTACAGAGGCACTTGAGACAGTTGACAACTCTTAAA	204				
Db	163	gtggataagctggaaactgtctcttcacgaggagcttgagacagttgacataactcttaa	222				
QY	205	GGCATCCTTGACAAACTGAAAGTGTACACTAGGAGTGCTTCAGAAATCCAGTCTTGCAA	264				
Db	223	ggcatctcttgaaacttgaaaggtcgacctaggagtgcttccagaaaacctagctcgacaa	282				
QY	265	CTGGCCAGCAAGAGGCCAGGAAGCTTGAGAAATTTGCTGAACAACTGCACTTTCTAAGCTG	324				
Db	283	ctggccaagcagaagggcccgaggagctgagaaattgtctgacaactgcaattcttaagctg	342				
QY	325	CTTCCAACTAACACGGACATTTTGGGTTGAAATCAGCAACTCCCTCATCTCTGGATGTC	384				
Db	343	cttccaactaaacacggaatcttttgggttgaaatacagcaactccctcatctctggaatgc	402				
QY	385	AAAGCTGAACCCATCGATGATGGCAAGAGCCCTTAACCTGAGCTTCCTGTCAACCCGGAAT	444				
Db	403	aaagctgaaccgatcgatgattggccaagaccttaacctgaagcttccctgtccaccgcgaat	462				
QY	445	GTCACTGTGGCGCGGCCCATCATTTGGGCCAGATTATCAACCTGAAGACCTCTTGGACCTTC	504				
Db	463	gtcaactgtggcgggcccatcatcttgccagattatcaacctgaaagccctctcttggaactc	522				
QY	515	CTGACGGCAGTCACAATTGAACACTGATCCCCAGACACACACGAGCTGTGCGGCTCTGGGA	564				
Db	523	ctgacggcagtcacaattgaaactgactccccagacacacacgagctgttgccgtctctggga	582				

QY	565	GAATGGCCAGTGAACCCACACAGCATCTCACATTTCTCTGCTGGACAAACACAGCCAAATC	624
Db	583	gaatggccagtgaccccaaccagcattctcatttctctgtggacaaacacagccaaatc	642
QY	625	ATCAACAAGTTCCTGAATAGCTGTATCAACAGCTGAAAAAGCACTGTATCTCTCCCTGCTG	684
Db	643	atcaacaagltctggatcagtcagtcgatacaacagctgaaaaagcaactgtatctctccctcgtg	702
QY	685	CAGAAGGAGATATGTCCATGTATCGCATCTTCATCTCCACTCCCTGGATGTGAATGTCAATT	744
Db	703	cagaaggagatgtccactgatcccatgattctccactcctcgtgatgtgaatgtcatt	762
QY	745	CACGAGTCTCTGCATTAATCCTCAGCACAAAACCCAGCTGCAAAACCTCATTTTGAAGAGGA	804
Db	763	cagcaggtcgtcgataatcctcagcaaaaaacccagctgcaaacacctcatctgaagagga	822
QY	805	CGAATGAGGAGGACACTGTGTGTGCATGCTGATTTGGTTCCAGTGGCTTGCCTCACACCCCC	864
Db	823	cgatgaggaggaccactgtgtgcatactgattgattggttccacagtggctgtcccaccccc	882
QY	865	TTATAGCATCTCCCTCCAGGAAGCTGCTGCCACCACTTAACACAGCGTGAAGACGCTGAGTC	924
Db	883	ttatagcatctcctccaggaaagctgtgcaccaccttaaccagcgtgaaagcctgagtc	942
QY	925	CCACCAGAAGACCTTCCAGATACCCCTTCTCTCAGTCAGAACACAGACGCTCTTACA	984
Db	943	ccaccagaaggaccttcccagataaccccttctccacagtcagaaacagcagcctctaca	1002
QY	985	CATGTTGTCTGCCCTCGG	1003
Db	1003	catgttctcatgacgaagg	1021

```

RESULT 12
US-08-749-388-2
; Sequence 2, Application US/08749288
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PAROTID
; TITLE OF INVENTION: SECRETORY PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08749,288
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0153 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-835-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```


1 TITLE OF INVENTION: POLYMORPHISMS IDENTIFIED THEREBY
2 FILE REFERENCE: GX-0020 P
3 CURRENT APPLICATION NUMBER: US/60/250,830
4 CURRENT FILING DATE: 2000-11-04
5 NUMBER OF SEQ ID NOS: 3246
6 SOFTWARE: PERL Program
7 SEQ ID NO 2204
8 LENGTH: 501
9 TYPE: DNA
10 ORGANISM: Homo sapiens
11 FEATURE:
12 NAME/KEY: misc_feature
13 OTHER INFORMATION: Incyte ID No: 903509.9
14 US-60-250-830-2204

Query Match 48.6% Score 499.4; DB 64; Length 501;
Best Local Similarity 99.8%; Pred. No. 3.8e-130;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 523 GAACTGATCCCGACACACACAGCCCTGTTGCCGCTGCGGAGAAATGCGCCAGTGACCCA 582
Db 1 gaaactgatccccagacacacagccctgttgcgctcctggagaatgcccagtgaccca 60
QY 583 ACCAGCATCTCACATTTCTTCTGCTGGAGAAACACAGCCCAATCATCAACAAGTTTCGTGAAT 642
Db 61 accagcatctcacatcttcttctgctggagaaacacagcccaatcatcaacaagtttcgtgaat 120
QY 643 ACCGTGATCAACAGCGCTGAAAGCACTGTATCTCCCTGCTGAGAGAGAGATATGTCCA 702
Db 121 agcgtgatcaacagcgtgaaagcaactgtatctccctgctgagagagagatatgtcca 180
QY 703 CTGATCCCGCATCTTCATCCACTCCCTGATGTGAATGTCTCATTGAGAGAGAGATATGTCCA 762
Db 181 ctgattcccgcatcttcattccactccctgattgtgaatgtctcattgagagagagatatgtcca 240
QY 763 CCTGAGCACAAACCCAGCTGCAACCCCTCATTTTGAAGAGAGAGAGATATGTCCA 822
Db 241 cctcagcaaaaacccagctgcaaacctcatctgaagagagagagatatgtcca 300
QY 823 GTGGTGATGCTGATTTGTTCCCACTGCTGAGTGAATGTCTCATTGAGAGAGAGATATGTCCA 882
Db 301 gtggtgatgctgatttggttcccaactgctgagtgaaatgtctcattgagagagagatatgtcca 360
QY 883 GGAAGCTGCTGCCACCACTAACAGCGTGAAGCCCTGATGCCACCCCTTATAGCATCTCCCTCCA 942
Db 361 ggaagctgctgccaccaactaacagcgtgaaagcctgagtcaccaaccagagagaccttcc 420
QY 943 CAGATACCCCTTCTCCTCACAGTCAAGACAGCCCTCTACACATGTGCTGCCCCCTG 1002
Db 421 cagatacccccttctcctcacagtcagaacagcagcctctacacatgtgtcctgccccctg 480
QY 1003 GCAATAAGGCCCATTTCTGC 1023
Db 481 gcaataaaggcccatcttctgc 501

RESULT 15

US-60-323-966-2204
Sequence 2204, Application US/60323966
GENERAL INFORMATION:
APPLICANT: Morris, MacDonald
APPLICANT: Lal, Preeti
APPLICANT: Diep, Dinh
TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
TITLE OF INVENTION: POLYMORPHISMS IDENTIFIED THEREBY
FILE REFERENCE: GX-0020-1 P
CURRENT APPLICATION NUMBER: US/60/323,966
CURRENT FILING DATE: 2001-09-21
NUMBER OF SEQ ID NOS: 3246
SOFTWARE: PERL Program
SEQ ID NO 2204

1 LENGTH: 501
2 TYPE: DNA
3 ORGANISM: Homo sapiens
4 FEATURE:
5 NAME/KEY: misc_feature
6 OTHER INFORMATION: Incyte ID No: 903509.9
7 US-60-323-966-2204

Query Match 48.6% Score 499.4; DB 71; Length 501;
Best Local Similarity 99.8%; Pred. No. 3.8e-130;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 523 GAACTGATCCCGACACACACAGCCCTGTTGCCGCTGCGGAGAAATGCGCCAGTGACCCA 582
Db 1 gaaactgatccccagacacacagccctgttgcgctcctggagaatgcccagtgaccca 60
QY 583 ACCAGCATCTCACATTTCTTCTGCTGGAGAAACACAGCCCAATCATCAACAAGTTTCGTGAAT 642
Db 61 accagcatctcacatcttcttctgctggagaaacacagcccaatcatcaacaagtttcgtgaat 120
QY 643 ACCGTGATCAACAGCGCTGAAAGCACTGTATCTCCCTGCTGAGAGAGAGATATGTCCA 702
Db 121 agcgtgatcaacagcgtgaaagcaactgtatctccctgctgagagagagatatgtcca 180
QY 703 CTGATCCCGCATCTTCATCCACTCCCTGATGTGAATGTCTCATTGAGAGAGAGATATGTCCA 762
Db 181 ctgattcccgcatcttcattccactccctgattgtgaatgtctcattgagagagagatatgtcca 240
QY 763 CCTGAGCACAAACCCAGCTGCAACCCCTCATTTTGAAGAGAGAGAGATATGTCCA 822
Db 241 cctcagcaaaaacccagctgcaaacctcatctgaagagagagagatatgtcca 300
QY 823 GTGGTGATGCTGATTTGTTCCCACTGCTGAGTGAATGTCTCATTGAGAGAGAGATATGTCCA 882
Db 301 gtggtgatgctgatttggttcccaactgctgagtgaaatgtctcattgagagagagatatgtcca 360
QY 883 GGAAGCTGCTGCCACCACTAACAGCGTGAAGCCCTGATGCCACCCCTTATAGCATCTCCCTCCA 942
Db 361 ggaagctgctgccaccaactaacagcgtgaaagcctgagtcaccaaccagagagaccttcc 420
QY 943 CAGATACCCCTTCTCCTCACAGTCAAGACAGCCCTCTACACATGTGCTGCCCCCTG 1002
Db 421 cagatacccccttctcctcacagtcagaacagcagcctctacacatgtgtcctgccccctg 480
QY 1003 GCAATAAGGCCCATTTCTGC 1023
Db 481 gcaataaaggcccatcttctgc 501

Search completed: August 6, 2002, 18:16:39
Job time: 6681 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 16:52:46 ; Search time 219.26 Seconds
(without alignments)
10202.412 Million cell updates/sec

Title: US-10-020-139-1
Perfect score: 1028
Sequence: 1 CACGACATTTCGACCATC.....AAGGCCCATTTCTGCAAAA 1028

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1362792 seqs, 1088025756 residues

Total number of hits satisfying chosen parameters: 2725584

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1028	100.0	1028	7	US-10-020-139-1
2	990.4	96.3	1049	7	US-10-119-480-159
3	383.8	37.3	449	7	US-10-020-139-10
4	374.8	36.5	538	7	US-10-020-139-11
5	312.8	30.4	359	7	US-10-020-139-12
6	303.6	29.5	374	7	US-10-020-139-13
7	258.8	25.2	406	7	US-10-020-139-14
8	229.8	22.4	493	7	US-10-020-139-15
9	129	12.5	395	7	US-10-020-139-16
10	103.2	10.0	116	7	US-10-020-139-17
11	71	6.9	756	7	US-10-027-632-11686
12	71	6.9	756	7	US-10-027-632-11687
13	41.6	4.0	360	7	US-10-020-139-18
14	41	4.0	999	7	US-10-176-912-434
15	41	4.0	999	7	US-10-179-524-434
16	41	4.0	999	7	US-10-184-634-434
17	41	4.0	999	7	US-10-184-644-434
18	40.4	3.9	1009	6	US-10-192-254-8
19	40.4	3.9	1017	6	US-10-192-254-9
20	40.4	3.9	1021	6	US-10-184-638-373
21	40.4	3.9	1021	6	US-10-187-594-373
22	40.4	3.9	1021	6	US-10-187-743-373
23	40.4	3.9	1021	6	US-10-187-748-373
24	40.4	3.9	1021	6	US-10-187-757-373
25	40.4	3.9	1021	6	US-10-187-589-373

26 40.4 3.9 1021 6 US-10-187-590-373 Sequence 373, App
27 40.4 3.9 1021 6 US-10-187-591-373 Sequence 373, App
28 40.4 3.9 1021 6 US-10-187-600-373 Sequence 373, App
29 40.4 3.9 1021 6 US-10-187-601-373 Sequence 373, App
30 40.4 3.9 1021 6 US-10-187-746-373 Sequence 373, App
31 40.4 3.9 1021 6 US-10-187-751-373 Sequence 373, App
32 40.4 3.9 1021 6 US-10-187-752-373 Sequence 373, App
33 40.4 3.9 1021 6 US-10-187-753-373 Sequence 373, App
34 40.4 3.9 1021 6 US-10-187-754-373 Sequence 373, App
35 40.4 3.9 1021 6 US-10-187-756-373 Sequence 373, App
36 40.4 3.9 1021 6 US-10-187-885-373 Sequence 373, App
37 40.4 3.9 1021 6 US-10-187-886-373 Sequence 373, App
38 40.4 3.9 1021 6 US-10-187-588-373 Sequence 373, App
39 40.4 3.9 1021 6 US-10-187-597-373 Sequence 373, App
40 40.4 3.9 1021 6 US-10-187-602-373 Sequence 373, App
41 40.4 3.9 1021 6 US-10-187-884-373 Sequence 373, App
42 40.4 3.9 1021 6 US-10-188-775-373 Sequence 373, App
43 40.4 3.9 1021 6 US-10-187-593-373 Sequence 373, App
44 40.4 3.9 1021 6 US-10-187-595-373 Sequence 373, App
45 40.4 3.9 1021 6 US-10-187-738-373 Sequence 373, App

ALIGNMENTS

RESULT 1
US-10-020-139-1
Sequence 1, Application US/10020139
GENERAL INFORMATION:
APPLICANT: DUAN, ROXANNE
RUBEN, STEVEN
TITLE OF INVENTION: Parotid Secretory Protein
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MD
COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/020.139
FILING DATE: 18-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/993.529
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF348
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 301-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1028 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 49..795
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 49..100
FEATURE:

NAME/KEY: mat_peptide
LOCATION: 103..795
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-020-139-1

Query Match 100.0%; Score 1028; DB 7; Length 1028;
Best Local Similarity 100.0%; Pred. No. 9.8e-281;
Matches 1028; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CACGAGATTTCAAGAGCATCTCTCTAAAGCGGTGTCAGACAAAGAGATGCTTCAGCTT 60
DB 1 CACGAGATTTCAAGAGCATCTCTCTAAAGCGGTGTCAGACAAAGAGATGCTTCAGCTT 60

QY 61 TCGAAACTGTTTCNCCTGTCGGCGGTGCTCACTGAGACCTCAGAGTCTCTTGTGCAAT 120
DB 61 TCGAAACTGTTTCNCCTGTCGGCGGTGCTCACTGAGACCTCAGAGTCTCTTGTGCAAT 120

QY 121 CTTGGCAATGACCTTAAGCAATGCTGGATAGCTGGAACCTGTTCTTCAAGGAGGACTT 180
DB 121 CTTGGCAATGACCTTAAGCAATGCTGGATAGCTGGAACCTGTTCTTCAAGGAGGACTT 180

QY 181 GAGACGTTGACATCTCTTAAGGCATCTTTGAGAACTGAAGTCAAGCTAGAGTG 240
DB 181 GAGACGTTGACATCTCTTAAGGCATCTTTGAGAACTGAAGTCAAGCTAGAGTG 240

QY 241 CTTGAGAAATCCAGTCTTGGCAACTGGCCAGCAGAGGCCAGGAAGCTGAGAAATG 300
DB 241 CTTGAGAAATCCAGTCTTGGCAACTGGCCAGCAGAGGCCAGGAAGCTGAGAAATG 300

QY 301 CTGAGCAATGCTATTTCTAGCTGCTTCCAACTAAGCAGGACATTTTGGGTGAAATC 360
DB 301 CTGAGCAATGCTATTTCTAGCTGCTTCCAACTAAGCAGGACATTTTGGGTGAAATC 360

QY 361 AGCAACTCCCTCATCTCTGATGTCGAAGCTGAACCGATGATGGCAAGGCCCTTAAC 420
DB 361 AGCAACTCCCTCATCTCTGATGTCGAAGCTGAACCGATGATGGCAAGGCCCTTAAC 420

QY 421 CTGAGCTTCCCTGTCACCGGAAATGTCAGTGGCGGGGCCATCATTTGGCCAGATTATC 480
DB 421 CTGAGCTTCCCTGTCACCGGAAATGTCAGTGGCGGGGCCATCATTTGGCCAGATTATC 480

QY 481 AAGCTGAAAGCCCTCTTGGACCTCTTACCGGAGTCACAAATTTGAAACTGATCCCGAGACA 540
DB 481 AAGCTGAAAGCCCTCTTGGACCTCTTACCGGAGTCACAAATTTGAAACTGATCCCGAGACA 540

QY 541 CACGAGCTTGTCCCTGCTGGGAGATGCGGAGTGACCCAGCCAGCAATCTCATCTTCC 600
DB 541 CACGAGCTTGTCCCTGCTGGGAGATGCGGAGTGACCCAGCCAGCAATCTCATCTTCC 600

QY 601 TTGCTGACAAACACAGCCAAATCATCAACAGTTCCGTAATAGCTGATCAACAGCTG 660
DB 601 TTGCTGACAAACACAGCCAAATCATCAACAGTTCCGTAATAGCTGATCAACAGCTG 660

QY 661 AAAGCACTGTATCTTCCCTGCTGCGAGAGGAGATATGTCCACTGATCCGATCTTCATC 720
DB 661 AAAGCACTGTATCTTCCCTGCTGCGAGAGGAGATATGTCCACTGATCCGATCTTCATC 720

QY 721 CACTCCCTGATGTGAATGCTATTCAGCAGCTGTCGATATTCCTCAGCAACAAACCCAG 780
DB 721 CACTCCCTGATGTGAATGCTATTCAGCAGCTGTCGATATTCCTCAGCAACAAACCCAG 780

QY 781 CTGCAACCTCTATTTGAAGAGAGCAATGAGGAGGACCACTGTGGTGCATGCTCATTTGG 840
DB 781 CTGCAACCTCTATTTGAAGAGAGCAATGAGGAGGACCACTGTGGTGCATGCTCATTTGG 840

QY 841 TTCCCACTGCTTGGCCCAACCCCTTATAGCATCTCCCTCAGGAAGCTGCTGCCACAC 900
DB 841 TTCCCACTGCTTGGCCCAACCCCTTATAGCATCTCCCTCAGGAAGCTGCTGCCACAC 900

QY 901 CTAACAGCGGTGAAGCCTGATGCCACAGAGGACCTTCCAGATACCCCTTCTCCTC 960
DB 901 CTAACAGCGGTGAAGCCTGATGCCACAGAGGACCTTCCAGATACCCCTTCTCCTC 960

QY 961 ACAGTCAGAACAGCAGCCTCTACACATGTTGCTGCTGCCCTGCAATAAAGGCCATTTC 1020
DB 961 ACAGTCAGAACAGCAGCCTCTACACATGTTGCTGCTGCCCTGCAATAAAGGCCATTTC 1020

QY 1021 TCGCAAAA 1028
DB 1021 TCGCAAAA 1028

RESULT 2
US-10-119-480-159
; Sequence 159, Application US/10119480
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc Gerriksen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3530P1C1
; CURRENT APPLICATION NUMBER: US/10/119,480
; CURRENT FILING DATE: 2002-04-09
; NUMBER OF SEQ ID NOS: 246
; Prior Application removed - See File Wrapper or Palm
; SEQ ID NO 159
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-119-480-159

Query Match 96.3%; Score 990.4; DB 7; Length 1049;
Best Local Similarity 99.4%; Pred. No. 4.3e-270;
Matches 994; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 25 TCTAAAGCGGTGTCAGACAAAGATGCTTCAGCTTTGGAACCTGTTCTCTGTGCGGC 84
DB 43 TCTGAGGTGTCAGACAAAGATGCTCAGCTTTGGAACCTGTTCTCTGTGCGGC 102

QY 85 GTGCTCACTGGGACCTCAGAGTCTCTTCTTGACAACTTTGGCAATGACCTTAAGCAATGTC 144
DB 103 GTGCTCACTGGGACCTCAGAGTCTCTTCTTGACAACTTTGGCAATGACCTTAAGCAATGTC 162

QY 145 GTGATGAAGTCTGAACTGTTCTTTCAGAGGAGTTCAGACAGTTCACAAATCTCTTAA 204
DB 163 GTGATGAAGTCTGAACTGTTCTTTCAGAGGAGTTCAGACAGTTCACAAATCTCTTAA 222

QY 205 GGCATCCTTGAGAACTGAAGTTCGACCTAGGAGTCTTCAGAAATCCAGTCTTGGCAA 264
DB 223 GGCATCCTTGAGAACTGAAGTTCGACCTAGGAGTCTTCAGAAATCCAGTCTTGGCAA 282

QY 265 CTGGCAAGCAGAGGCCCGAGGAGTTCGAAATGCTGAACTGCTCATTTCTAAGCTG 324
DB 283 CTGGCAAGCAGAGGCCCGAGGAGTTCGAAATGCTGAACTGCTCATTTCTAAGCTG 342

QY 325 CTTCACCACTAAGCAGGACATTTTGGGTGAAATCAGCAACTCCCTCATCTCGATGTC 384
DB 343 CTTCACCACTAAGCAGGACATTTTGGGTGAAATCAGCAACTCCCTCATCTCGATGTC 402

QY 385 AAGCTGAACCGATCGATGATGGCAAGCCCTTAACCTGAGCTTCCCTGTGTCACGCGCAAT 444
DB 403 AAGCTGAACCGATCGATGATGGCAAGCCCTTAACCTGAGCTTCCCTGTGTCACGCGCAAT 462

QY 445 GTCACTGTGGCGGCCCATCATTTGGCCAGATTATCAACCTGAAAGCCTCTTGGACCTC 504
DB 463 GTCACTGTGGCGGCCCATCATTTGGCCAGATTATCAACCTGAAAGCCTCTTGGACCTC 522

```

QY 505 CTGACGCGCTCACAATTGAAATGATCCCGACACACACCGCTGTTGCGCTCTCGGA 564
    |||
Db 523 ctgacgcagctcaaatgaaactgattccagacacacacacccctgttgccctcgga 582
    |||
QY 565 GAATGCGCCAGTGAACCAACAGCATCTCAGTTCTTCTGCAACAAACAGCAAAATC 624
    |||
Db 583 gaatgcccagtgaccacaacagcatctcacttctctgtggaacacacagccaaatc 642
    |||
QY 625 ATCAACAAGTTCTGATAGCGTGCATCAACAGCTGAAAGACACTGATCTCCCTGCTG 684
    |||
Db 643 atcaacaagttctgataagcgtgatacaacacgctgaaagacactgtatccccctgtg 702
    |||
QY 685 CACAAGGAGATATGTCACACTGATCCGATCTTCATCCACTCCCTGATGTGAATCAT 744
    |||
Db 703 cagaagagagalaagtccactgattccgcatcttcacacacccctgagtgatgcaatt 762
    |||
QY 745 CACGAGTCTGCTGATATCTCAGCAACAAACCCAGCTCAACCCCTATTTGAAGAGA 804
    |||
Db 763 cagcaggtcgtgataatccctcagcaaaaacccagctgcaaacctctcatctgaagagga 822
    |||
QY 805 CGAATGAGGAGGACCACTGTGTGTCATGTGATTTGCTTCCAGTGGCTTGGCCACCC 864
    |||
Db 823 cgaatgaggagagaccactgtgtgcatgtgtgtgttccagtggtgtgcccaccccc 882
    |||
QY 865 TTATAGCATCTCCCTCCAGGAGCTGTCGCCACCACTAACCCAGGTGAAGCCTGAGTC 924
    |||
Db 883 ttatagcatctccctccaggaagctgtgtccacacacccactaacagcgtgaaagctgagtc 942
    |||
QY 925 CCACGAGAAGGACCTTCCAGATACCCCTTCTCTCCACAGTCAAGACAGCAGCCTTACA 984
    |||
Db 943 ccacgagaagaccttccagataaaccttctctcacagtcagaaacagcagcctctaca 1002
    |||
QY 985 CATCTGTCTGCTGCCCTGGCAATAAAGGCCATTTCGTGCA 1024
    |||
Db 1003 catgtgtctgctgccctggcaataaaggccatttctgca 1042
    |||

```

RESULT 3

```

US-10-020-139-10
; Sequence 10, Application US/10020139
; GENERAL INFORMATION:
; APPLICANT: DUAN, ROXANNE
; TITLE OF INVENTION: Parotid Secretory Protein
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/020,139
; FILING DATE: 18-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,529
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF348
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 301-8439
; INFORMATION FOR SEQ ID NO: 10:

```

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 449 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-020-139-10

```

```

Query Match 37.3%; Score 383.8; DB 7; Length 449;
Best Local Similarity 95.5%; Pred. No. 1.6e-98;
Matches 427; Conservative 0; Mismatches 11; Indels 9; Gaps 3:

```

```

QY 1 CACGAGATTTTCATGAGCATCTCTCTAAACGGGTGTCGAAGCAAAAGATGCTTCAGCTT 60
    |||
Db 3 CACGAGATTTTCATGAGCATCTCTCTAAACGGGTGTCGAAGCAAAAGATGCTTCAGCTT 62
    |||
QY 61 TGGAACTTTCTCTCTGTCGGCGCTGCTCACTGGGACCTCAGAGTCTCTTCTTACAAAT 120
    |||
Db 63 TGGAACTTTCTCTCTGTCGGCGTCTCACTGGGACCTCAGAGTCTCTTCTTACAAAT 122
    |||
QY 121 CTGGCAATCACCTTAAGCAATGTCGTGGATAAGCTGGAACCTGTTCTTCCAGAGGACTT 180
    |||
Db 123 CTGGCAATCACCTTAAGCAATGTCGTGGATAAGCTGGAACCTGTTCTTCCAGAGGACTT 182
    |||
QY 181 GAGACATTTGACATCTCTTAAAGG-----CATCTTGGAGAACTGAAAGTGCACCT 233
    |||
Db 183 GAGACATTTGACATCTCTTAAAGGGCATCCCTTTTNGAGAACTGAAAGTGCACCT 242
    |||
QY 234 AGGAGTCTTTCAGAAATCCAGTCTTGGCAACTGGCCAGCAAGCAAGGCCAGGAAGTCA 293
    |||
Db 243 AGGAGTCTTTCAGAAATCCAGTCTTGGCAACTGGCCAGCAAGCAAGGCCAGGAAGTCA 302
    |||
QY 294 GAAATGCTGAA--CAATGTCAATTTCTAAGCTGCTTCCAACTAACACGGACATTTTGGGT 352
    |||
Db 303 GAAATGCTGAA--CAATGTCAATTTCTAAGCTGCTTCCAACTAACACGGACATTTTGGGT 362
    |||
QY 353 TG-AAATCAGCAACTCCCTCATCTCTGGATGTCAAAGCTGAACCGCATCATGATGCAAA 411
    |||
Db 363 TGAAATCAGCAACTCCCTCATCTCTGGATGTCAAAGCTGAACCGCATCATGATGCAAA 422
    |||
QY 412 GGCCTTAACCTGAGCTTCCCTGTCAACC 438
    |||
Db 423 GGCCTTAACCTGAGCTTCCCTGTCAACC 449
    |||

```

RESULT 4

```

US-10-020-139-11
; Sequence 11, Application US/10020139
; GENERAL INFORMATION:
; APPLICANT: DUAN, ROXANNE
; TITLE OF INVENTION: Parotid Secretory Protein
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/020,139
; FILING DATE: 18-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,529

```

```

COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/020,139
FILING DATE: 18-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/993,529
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF348
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 301-8439
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-020-139-12

Query Match          30.4%;   Score 312.8;   DB 7;   Length 359;
Best Local Similarity 97.2%;   Pred No. 1.8e-78;
Matches 347;   Conservative 0;   Mismatches 7;   Indels 3;   Gaps

QY 1 CACGAGATTTCATGAGCATCCTCTCTAAACGCGTGTCAAGACAAAAGATGCTTCAGCTT 60
DB 3 CACGAGATTTCATGAGCATCCTCTCTAAACGCGTGTCAAGACAAAAGATGCTTCAGCTT 62
QY 61 TGGAACTTGTTCTCCGTGCGCGTGTCTACTGGGACCTCAGAGTCTCTCTTTGACAAT 120
DB 63 TGGAACTTGTTCTCCGTGCGCGTGTCTACTGGGACCTCAGAGTCTCTCTTTGACAAT 122
QY 121 CTTGGCAATGACCTAAGCAATGTCTGGTGAATAGCT- GGAACCTGTTTTCACGAGGACT 179
DB 123 CTTGGCAATGACCTAAGCAATGTCTGGTGAATAGCTTGGGAACCTGTTTTCACGAGGACT 182
QY 180 TCACACAGCTTGACAATACTCTTAAAGGCATCTCTTCAGAAACTGAAGTCACCTTAGAGT 239
DB 183 TCACACAGCTTGACAATACTCTTAAAGGCATCTCTTCAGAAACTGAAGTCACCTTAGAGT 242
QY 240 GCTTCAGAAATCCAGTCTTGGCAACTGGCCCAAGCAGAAGG- CCCAGAGACTGAGAAAT 298
DB 243 GCTTCAGAAATCCAGTCTTGGCAACTGGNCAGCAGAAGGCCAGGAAGCTGAGAAAT 302
QY 299 TGCT- GAACAATGTCAATTTCTAAGCTGCTTCCAACCTAACACGGACATTTTGGGTTG 354
DB 303 TCCTGCANCAATGTCANCTCTAAGCTGNTTCCGACTACACGNCATTTTGGGTTG 359

RESULT 6
US-10-020-139-13
; Sequence 13, Application US/10020139
; GENERAL INFORMATION:
; APPLICANT: DUAN, ROXANNE
; TITLE OF INVENTION: Parolid Secretory Protein
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE

```

```

?
? STATE: MD
?
? COUNTRY: US
?
? ZIP: 20850
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Floppy disk
?
? COMPUTER: IBM PC compatible
?
? OPERATING SYSTEM: PC-DOS/MS-DOS
?
? SOFTWARE: Patent In Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
?
? APPLICATION NUMBER: US/10/020,139
?
? FILING DATE: 18-Dec-2001
?
? CLASSIFICATION: <Unknown>
?
? PRIOR APPLICATION DATA:
?
? APPLICATION NUMBER: US/08/993,529
?
? FILING DATE: <Unknown>
?
? ATTORNEY/AGENT INFORMATION:
?
? NAME: BROOKES, ANDERS A.
?
? REGISTRATION NUMBER: 36,373
?
? REFERENCE/DOCKET NUMBER: PF348
?
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: (301) 309-8504
?
? TELEFAX: (301) 301-8439
?
? INFORMATION FOR SEQ ID NO: 13:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 374 base pairs
?
? TYPE: nucleic acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: DNA (genomic)
?
? SEQUENCE DESCRIPTION: SEQ ID NO: 13:
?
? US-10-020-139-13

```

Query Match	29.5%	Score 303.6	DB 7	Length 374
Best Local Similarity	95.0%	Pred. No. 7.5e-76		
Matches 343	Conservative 0	Mismatches 14	Indels 4	Gaps
QY 644	GGGTGATCAACAGCCTGAAGGACCTGTATCTCTCTCTGCTGCAGAGGAGATATGTCAC	703		
Db 14	GGGTGATCAACAGCCTGAAGGACCTGTATCTCTCTCTGCTGCAGAGGAGATATGTCAC	73		
QY 704	TGATCCGATCTTTCATCCACTCCCTCGATGTGAATGTATTCCACAGCTCTGCTCAATC	763		
Db 74	TGATCCGATCTTTCATCCACTCCCTCGATGTGAATGTATTCCACAGCTCTGCTCAATC	133		
QY 764	CTCAGCAAAACCCAGCTGCAAAACCTCATTTGAAGAGGAGGAATGAGGAGGACCACTG	823		
Db 134	CTCAGCAAAACCCAGCTGCAAAACCTCATCTGAAGAGGACGAATGAGGAGGACCACTG	193		
QY 824	TGGTGCATGCTGATTGGTTGCCAGTGGCTTGCCCAACCCCTTATAGCATCTCCCTCCAG	883		
Db 194	TGGTGCATGCTGATTGGTTGCCAGTGGCTTGCCCAACCCCTTATAGCATCTCCCTCCAG	253		
QY 884	GAAGTGTGTCACACCTTACCAGCGTGAAGCCTTGAGTCCCAACAGAGGACCTTCC	942		
Db 254	GAAGTGTGTCACACCTTACCAGCGTGAAGCCTTGAGTCCCAACAGAGGACCTTCC	313		
QY 943	CAGATACCCC-TTCTCTCTCAGTTCAGACAGCAG--CCTCTACACATGTTGTCCTGCC	999		
Db 314	CAGATACCCCTTTTCTCTCAGTTCAGAGGNGNGGCTCTTACACNTGTTGTCCGCGCC	373		
QY 1000	C 1000			
Db 374	C 374			

```

RESULT      7
US-10-020-139-14
; Sequence 14, Application US/10020139
; GENERAL INFORMATION:
; APPLICANT: DUAN, ROYANNE
;
; APPLICANT: RUBEN, STEVEN
;
; TITLE OF INVENTION: Parotid S

```

NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: HUMAN GENOME SCIENCES, INC.
 STREET: 9410 KEY WEST AVENUE
 CITY: ROCKVILLE
 STATE: MD
 COUNTRY: US
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/020,139
 FILING DATE: 18-Dec-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/993,529
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: BROOKES, ANDERS A.
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PF348
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 301-8439
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 406 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 14:
 US-10-020-139-14

Query Match	25.2%	Score 258.8;	DB 7;	Length 406;
Best Local Similarity	89.4%;	Pred. No. 3.7e-63;		
Matches 313;	Conservative	0;	Mismatches 29;	Indels 8;
				Caps
QY	9	TTCATGAGCATCTCTCTTAACCGCTGTCAAGACAAAAGATGCTTCAGCTTTGGAAACT	68	
Db	1	TTCATGAGCATCTCTCTTAACCGCTGTCAAGACAAAAGATGCTTCAGCTTTGGAAACT	59	
QY	69	TGTTCTCTCTGTGGCGGTGCTCACTGGAGCCTCAGAGTCTCTCTCTTGACAATCTTGGCAA	128	
Db	60	TGTTCTCTCTGTGGCGGTGCTCACTGGAGCCTCAGAGTCTCTCTTGACAATCTTGGCAA	119	
QY	129	TGACCTAAGCAATGCTGGATAGCTGGAACGTGTCTTCACGAGGACCTTGAGACAGT	188	
Db	120	TGACCTAAGCAATGCTGGATAGCTGGAACGTGTCTTCACGAGGACCTTGAGACAGT	179	
QY	189	TGACAATACTCTTAAGGSCATCCTTCAGAACTTGAAGTTCGACCTAGGAGTGCTTCAGAA	248	
Db	180	TGACAATACTCTTAAGGSCATCCTTCAGAACTTGAAGTTCGACCTAGGAGTGCTTCAGAA	239	
QY	249	ATCCAGTGCTTTGGCAACTGGGCCAAGCAGAGGCC --- AGGAAGCTGAGAAATGGTG ---	303	
Db	240	ATCCAGTGCTTTGGCAACTGGGCCAAGCAGAGGCC --- AGGAAGCTGAGAAATGGTG ---	299	
QY	304	--AACAATGTCATTCTTAAGCTGCTTCCAACATAACACGGACATTTTTGGG	351	
Db	300	ACAAATGTTCAATTTCAAGCTGCTTTCCAACATAACACGGACATTTTTGGG	349	

RESULT 8
US-10-020-139-15
; Sequence 15, Application US/10020139
; GENERAL INFORMATION:
; APPLICANT: DUAN, ROXANNE
; RUBEN, STEVEN

; TITLE OF INVENTION: Parotid Secretory Protein
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
 ; STREET: 9410 KEY WEST AVENUE
 ; CITY: ROCKVILLE
 ; STATE: MD
 ; COUNTRY: US
 ; ZIP: 20850

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/10/020,139
 ; FILING DATE: 18-Dec-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/993,529
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BROOKES, ANDERS A.
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PF348
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 301-8439

; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 493 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
 ; US-10-020-139-15

Query Match 22.4%; Score 229.8; DB 7; Length 493;
 Best Local Similarity 93.4%; Pred. No. 6.3e-55;
 Matches 240; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 602 TGTGTGCAAAACACAGCCAAATCATCAACAAGTTCGTGAATAGCGTGATCAACACGGTGA 661
 Db 83 TGTGTTCACAGACAGACAGCCAAATCATCAACAAGTTCGTGAATAGCGTGATCAACACGGTGA 142
 QY 662 AAAGCACTGTATCTCCCTCCCTGCGAGAGAGATATGTCCTGATGTCGATCTTCATCC 721
 Db 143 AAAGCACTGTATCTCCCTGCGAGAGAGATATGTCCTGATGTCGATCTTCATCC 202
 QY 722 ACTCCCTGGATGTGAATGTCAATCAUCAGGTGTCGATTAATCTCAGCACAAAACCCAGC 781
 Db 203 ACTCCCTGGATGTGAATGTCAATCAUCAGGTGTCGATTAATCTCAGCACAAAACCCAGC 262
 QY 782 TGAACAACTCATTTCAAGAGGACGAATCAGAGGACCACTGTGGTGATGCTGATTCGT 841
 Db 263 TGAACAACTCATTTCAAGAGGACGAATCAGAGGACCACTGTGGTGATGCTGATTCGT 322
 QY 842 TCCCAAGTGGCTTCCGCC 858
 Db 323 AGCCAGTCTGTGTGCC 339

RESULT 9
 US-10-020-139-16
 ; Sequence 16, Application US/10020139
 ; GENERAL INFORMATION:
 ; APPLICANT: DUAN, ROXANNE
 ; RUBEN, STEVEN
 ; TITLE OF INVENTION: Parotid Secretory Protein
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
 ; STREET: 9410 KEY WEST AVENUE
 ; CITY: ROCKVILLE
 ; STATE: MD
 ; COUNTRY: US
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/10/020,139
 ; FILING DATE: 18-Dec-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/993,529
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BROOKES, ANDERS A.
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PF348
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 301-8439
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 395 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
 ; US-10-020-139-16

Query Match 12.5%; Score 129; DB 7; Length 395;
 Best Local Similarity 76.6%; Pred. No. 2e-26;
 Matches 279; Conservative 0; Mismatches 65; Indels 20; Gaps 10;
 QY 685 CAGAAGGAGATATGTCCTGATCCGATCCGAT--CTTCATCCACTCCCTGGATGTGAATGT-- 740
 Db 26 CAGAAGGAGATATGTCCTGATCCGATCCGATCCCTGGATGTGAATGTTC 85
 QY 741 CATTCAGCAGTGT--CGATAATCT--CAGCACAAACCCAGCTGCAACCC--TCATTT 796
 Db 86 ATTNCAGCAGTGTCTGCGATATCCNGCAGCACAAACCCAGCTGCGAACCCTTCATCT 145
 QY 797 GAAGAGGACGAATCAGAGGAGCCACTGT--GGTGCATGCTGATTGTTGCC--AGTGGCT 852
 Db 146 GAAGAGGACGAATCAGAGGAGCCACTGTGGTGCATGCTGATTGTTGCCAGTGGCTT 205
 QY 853 TGCCCCACCCCTTATAGCATCTCCCTCCAGGAAGCTGTGCA--CCACCTAACCCAGCG 910
 Db 206 GCCCCACCCCTTANAGCANTCCCTCCAGGAAGCTGCTGNAACCAACCAACCCAGCG 265
 QY 911 TGAAGCCTGAGTCCACCAGGAGACCTTCCAGATACCCCTTCTCTCTCA--CAGTCAG 968
 Db 266 TGAAGCCTGATNCCACCAGAGGACCTTCCAGATACCCCTGCTNCNCAACAGTNAAG 325
 QY 969 AACAGCAGCTCT--CTACACATGTTCTCTGCTCCCTGGCATAAAGGCCCATTTCTGCA 1024
 Db 326 AACAGCAGCTCTGACACACATGNGTTCGTGGCCCCCGGCGCAATAAAAGGCCCATTTTGGC 385
 QY 1025 AAAA 1028
 Db 386 AAAA 389

RESULT 10
 US-10-020-139-17
 ; Sequence 17, Application US/10020139
 ; GENERAL INFORMATION:

1 APPLICANT: DUAN, ROXANNE
 2 RUBEN, STEVEN
 3 TITLE OF INVENTION: Parotid Secretory Protein
 4 NUMBER OF SEQUENCES: 18
 5 CORRESPONDENCE ADDRESS:
 6 ADDRESS: HUMAN GENOME SCIENCES, INC.
 7 STREET: 9410 KEY WEST AVENUE
 8 CITY: ROCKVILLE
 9 STATE: MD
 10 COUNTRY: US
 11 ZIP: 20850
 12
 13 COMPUTER READABLE FORM:
 14 MEDIUM TYPE: Floppy disk
 15 COMPUTER: IBM PC compatible
 16 OPERATING SYSTEM: PC-DOS/MS-DOS
 17 SOFTWARE: Patent In Release #1.0, Version #1.30
 18
 19 CURRENT APPLICATION DATA:
 20 APPLICATION NUMBER: US/10/020,139
 21 FILING DATE: 18-Dec-2001
 22 CLASSIFICATION: <Unknown>
 23
 24 PRIORITY APPLICATION DATA:
 25 APPLICATION NUMBER: US/08/993,529
 26 FILING DATE: <Unknown>
 27
 28 ATTORNEY/AGENT INFORMATION:
 29 NAME: BROOKES, ANDERS A.
 30 REGISTRATION NUMBER: 36,373
 31 REFERENCE/DOCKET NUMBER: PF348
 32 TELECOMMUNICATION INFORMATION:
 33 TELEPHONE: (301) 309-8504
 34 TELEFAX: (301) 301-8439
 35
 36 INFORMATION FOR SEQ ID NO: 17:
 37 SEQUENCE CHARACTERISTICS:
 38 LENGTH: 116 base pairs
 39 TYPE: nucleic acid
 40 STRANDEDNESS: single
 41 TOPOLOGY: linear
 42
 43 MOLECULE TYPE: DNA (genomic)
 44 SEQUENCE DESCRIPTION: SEQ ID NO: 17:
 45 AS-10-020-139-17

```

Query Match      10.08; Score 103.2; DB 7; Length 116;
Best Local Similarity 92.19; P-Val 2.7e-19;
Matches 105; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      1  CACGAGATTTCATGAGCATCTCTCTTAAACGGGTGTCAAGACAAAAGATGCTTCAGCTT 60
      |||||
Db       3  CACGANATTTCATGAGCATCTCTCTTAAACAGCGTGTCAAGACAAAAGATGCTTCAGCTT 62

QY      61  TCGAACAATGTTCTCCTGTGGGGGTCTCACTGGGACCTCAGAGTCTCTTCTT 114
      |||||
Db       63  TCGNACAATGTTCTCTATNCGCGTCTCACTGACCTCAGAAATCTCTNCCTT 116

RESULT      11
US-10-027-632-11686
: Sequence 11686, Application US/10027632
: GENERAL INFORMATION:
: APPLICANT: Wang, David G.
: TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
: POLYMORPHISMS IN THE HUMAN GENOME
: FILE REFERENCE: 108827.129
: CURRENT APPLICATION NUMBER: US/10/027,632
: CURRENT FILING DATE: 2002-04-30
: PRIOR APPLICATION NUMBER: US 60/218,006
: PRIOR FILING DATE: 2000-07-12
: PRIOR APPLICATION NUMBER: US 60/198,676
: PRIOR FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: US 60/193,483
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: US 60/185,218
: PRIOR FILING DATE: 2000-02-24
: PRIOR APPLICATION NUMBER: US 60/167,363

```

```

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11686
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11686

Query Match          6.9%; Score 71; DB 7; Length 756;
Best Local Similarity 86.5%; Pred. No. 6.3e-10;
Matches 77; Conservative 1; Mismatches 11; Indels 0; Gaps
    QY 673 TCCTCCCTGCTGCAGAGGAGATATGCCACTGATCGCGCATCTTCATCCACTCCCTCGAT 732
    Db 45 tctctcttccgctgcagatatgccactgatctgcattccatccactccctggat 104

    QY 733 GTGAATGTCATTCAGCAGGCTCGTCGATAA 761
    Db 105 gtgaatgcttcattcagcaggctcstcgtaa 133

RESULT 12
US-10-027-632-11687
; Sequence 11687, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11687
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11687

Query Match          6.9%; Score 71; DB 7; Length 756;
Best Local Similarity 86.5%; Pred. No. 6.3e-10;
Matches 77; Conservative 1; Mismatches 11; Indels 0; Gaps
    QY 673 TCCTCCCTGCTGCAGAGGAGATATGCCACTGATCGCGCATCTTCATCCACTCCCTCGAT 732
    Db 45 tctctcttccgctgcagatatgccactgatctgcattccatccactccctggat 104

    QY 733 GTGAATGTCATTCAGCAGGCTCGTCGATAA 761
    Db 105 gtgaatgcttcattcagcaggctcstcgtaa 133

```

```
Query Match      6.9%; Score 71; DB 7; Length 756;
Best Local Similarity 80.5%; Pred. NO. 6.3e-10;
Matches 77; Conservative 11; Mismatches 10; Indels 0; Gaps 0;
```

	QY	673	TCTCCCTGCTGCAGAGGAGATATGCCACTGATCGCGATCTTCATCCACCTCCTGGAT	732
Db	45	tctccttctccggtgcagatgatgccactgatctgcattcttcacacctccctggat	104	
QY	-	733	GTGAATGTCATTACAGACGGTCCGCGATAA	761
Db	105	qtgaatgtcattccagacggtcstccqqlaa	133	

Search completed: August 6, 2002, 18:20:30
Job time: 5264 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 16:10:18 ; Search time 50.07 Seconds
(without alignments)
5043.164 Million cell updates/sec

Title: US-10-020-139-1
Perfect score: 1028
Sequence: 1 CACGAGATTTCATGACGATC.....AAGGCCCATTTCTGCAAAA 1028

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767056

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.4	3.5	1215	2	US-09-092-770-8
2	36.4	3.5	1215	4	US-09-222-851-8
3	34.8	3.4	1214	2	US-09-092-770-7
4	34.8	3.4	1214	4	US-09-222-851-7
5	34	3.3	1870	3	US-09-235-186-9
6	33.8	3.3	44377	2	US-08-804-227C-7
7	33.8	3.3	44377	2	US-08-804-198-1
8	33.6	3.3	942	2	US-08-446-806-2
9	33.6	3.3	942	4	US-09-385-028-17
10	33.6	3.3	11604	4	US-09-385-028-13
11	33.6	3.3	15079	4	US-09-385-028-1
12	32.2	3.1	1956	3	US-08-693-940-2
13	32.2	3.1	2322	3	US-08-564-264-2
14	32.2	3.1	4963	1	US-08-076-726-16
15	32.2	3.1	4963	1	US-08-260-452-9
16	32.2	3.1	4963	2	US-08-481-970-9
17	32.2	3.1	4963	2	US-08-897-719-9
18	32.2	3.1	4963	4	US-09-163-269-9
19	32.2	3.1	6450	4	US-09-041-886-34
20	32	3.1	693	2	US-08-791-347-16
21	32	3.1	2040	2	US-08-533-669A-5
22	31.8	3.1	6953	1	US-07-805-123C-2
23	31.8	3.1	6953	1	US-08-033-081B-2
24	31.6	3.1	1212	2	US-09-092-770-18
25	31.6	3.1	1212	4	US-09-222-851-18
26	31.6	3.1	1294	3	US-09-025-691-2
27	31.4	3.1	196	5	PCT-US93-06251-15

C 28	31.2	3.0	3671	1	US-08-176-620A-5	Sequence 5, Appl
C 29	31.2	3.0	3671	1	US-08-463-862-5	Sequence 5, Appl
C 30	31.2	3.0	3671	2	US-08-461-985-5	Sequence 5, Appl
C 31	31.2	3.0	3671	2	US-08-458-887B-5	Sequence 5, Appl
C 32	31.2	3.0	3671	4	US-08-932-787B-5	Sequence 5, Appl
C 33	31.2	3.0	3671	4	US-08-932-012C-5	Sequence 5, Appl
C 34	31.2	3.0	3671	4	US-08-888-818C-5	Sequence 5, Appl
C 35	31	3.0	529	3	US-09-188-930-24	Sequence 24, Appl
C 36	31	3.0	529	3	US-09-188-930-200	Sequence 200, App
C 37	31	3.0	7521	4	US-09-004-838-116	Sequence 116, App
C 38	31	3.0	10815	4	US-09-004-838-21	Sequence 21, Appl
C 39	31	3.0	13149	4	US-09-004-838-87	Sequence 87, Appl
C 40	30.6	3.0	975	4	US-09-365-150-3	Sequence 3, Appl
C 41	30.6	3.0	975	4	US-09-365-150-4	Sequence 4, Appl
C 42	30.6	3.0	1114	3	US-09-048-889-10	Sequence 10, Appl
C 43	30.6	3.0	1769	1	US-08-253-155A-19	Sequence 19, Appl
C 44	30.6	3.0	2593	1	US-08-728-956-1	Sequence 1, Appl
C 45	30.6	3.0	3450	2	US-08-545-562A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-092-770-8
; Sequence 8, Application US/09092770
; Patent No. 5973119
; GENERAL INFORMATION:
; APPLICANT: Coats, Steven R.
; APPLICANT: Bass, Michael B.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5973119el Cyclin E Genes and Proteins
; FILE REFERENCE: A-524
; CURRENT APPLICATION NUMBER: US/09/092.770
; CURRENT FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Human
US-09-092-770-8

Query Match	3.5%	Score 36.4	DB 2	Length 1215
Best Local Similarity	25.9%	Pred. No. 0.15	183	Indels 0
Matches 85	Conservative 60	Mismatches 183	Indels 0	Gaps 0
QY	239	TGCTTCAGAAATCCAGTCTGGCAACTGGCCCAAGCAGAGAGCCGAGAGCTGAGAAAT	298	
Db	446	tnyngarglntgygargintayacnycaymgngaraenttytlaytngcncargayt	505	
QY	299	TGCTGAACAATGCTCATTTCTTAAGCTGCTTCCCAACTTAACACGAGACATTTTGGGTTGAAA	358	
Db	506	tyttgygmnttyatgynacncaraarayathaaraayatygtncarytnathg	565	
QY	359	TCAGCAACTTCCTCATCTCTGGATGTCAAGCTGAACCGATCGATGCGCAAGGCGCTTA	418	
Db	566	gnathacnwsnytnnttyathgcnwsnaarytngargarathtaygcnccnaarytncarg	625	
QY	419	ACCTGAGCTTCCTGTCACCCGGAATGTCACTGTGGCGGCGCCCATCATTTGGCCGAGATTA	478	
Db	626	arttygcntatgtnacnagaygngcntgywngargarayathytmgntatggarytna	685	
QY	479	TCACCTGAAAGCCCTTCCTTGGACCTCTGACCCGAGCAGTACAAATGAAACTGATCCCCAGA	538	
Db	686	thatytnaargcnytnaartgggarytntgyccngtnacnathatwsntggtytnaay	745	
QY	539	CACACACGCTGTTGCCCTCTCTGGGAGA	566	
Db	746	tnnttytncargtngaygcnynnaarga	773	


```

RESULT      6
US-08-804-227C-7
: Sequence 7, Application US/08804227C
: Patent No. 5876991
: GENERAL INFORMATION:
: APPLICANT: DeHoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rosteck, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:

```

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/385.028
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/790.462
 FILING DATE: 29-JAN-1997
 ATTORNEY/AGENT INFORMATION:


```

; NAME/KEY: misc_feature
; LOCATION: (1270-1284)
; OTHER INFORMATION: /note= "Linker peptide."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1285-2322)
; OTHER INFORMATION: /note= "Estrogen binding domain."
US-08-564-264-2

Query Match      3.1%; Score 32.2; DB 3; Length 2322;
Best Local Similarity 48.6%; Pred. No. 4.5;
Matches 88; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 583 ACCAGCATCTCATTCTTCTGCTGGACAAACACAGCCAAATCATCAACAAGTTCTGTGAAT 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1927 ACCCTGAAGTCTCTGGAGAGAGAGGACCATATCCACCGAGTCTGTGGACAAAGATCAACAGAC 1986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 643 AGCGTGATCAACAGCGCTGAAAGACACTGTATCTCTCCCTGCTGCAGAGGAGATATGTCCA 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1987 ACTTTGATCCACCTGATGGCCAGGAGCGCTGACCTCGAGCAGCAGCAGCGGCTG 2046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 703 CTGATCCGCATCTTCATCCACTCCCTCGATGTGAATGTCAATCAGCAGGTCTGTGATAAT 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2047 GCCAGCTCTCTCTCATCTCTCCCATCATCAGGCACATGAGTAACAAAGGCATGGAGCAT 2106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 763 C 763
Db 2107 C 2107

RESULT 14
US-08-076-726-16
; Sequence 16, Application US/08076726
; Patent No. 5464758
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Tight Control of Gene Expression in
; TITLE OF INVENTION: Eucaryotic Cells by Tetracycline-responsive Promoters
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,726
; FILING DATE: 14-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942.2490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 16:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
US-08-076-726-16

Query Match      3.1%; Score 32.2; DB 3; Length 2322;
Best Local Similarity 48.6%; Pred. No. 4.5;
Matches 88; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 583 ACCAGCATCTCATTCTTCTGCTGGACAAACACAGCCAAATCATCAACAAGTTCTGTGAAT 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1927 ACCCTGAAGTCTCTGGAGAGAGGACCATATCCACCGAGTCTGTGGACAAAGATCAACAGAC 1986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 643 AGCGTGATCAACAGCGCTGAAAGACACTGTATCTCTCCCTGCTGCAGAGGAGATATGTCCA 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1987 ACTTTGATCCACCTGATGGCCAGGAGCGCTGACCTCGAGCAGCAGCAGCGGCTG 2046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 703 CTGATCCGCATCTTCATCCACTCCCTCGATGTGAATGTCAATCAGCAGGTCTGTGATAAT 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1978 GCCAGCTCTCTCTCATCTCTCCCATCATCAGGCACATGAGTAACAAAGGCATGGAGCAT 2037
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 763 C 763
Db 2038 C 2038

RESULT 15
US-08-260-452-9
; Sequence 9, Application US/08260452
; Patent No. 5650298
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Saifeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
; TITLE OF INVENTION: Cells by Tetracycline-responsive Promoters
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lanhive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/260,452
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076,327
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-013CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; IMMEDIATE SOURCE:
; CLONE: pUHD BGR4
US-08-260-452-9
```

Search completed: August 6, 2002, 17:25:33
Job time: 4515 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:49:58 ; Search time 66.51 Seconds
(without alignments)
359,739 Million cell updates/sec

Title: US-10-020-139-2
Perfect score: 1233
Sequence: 1 MLQWLKLVLLCGVLTGTSSES.....NVIQVVVDNPNQHKTLQTLI 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	364.5	29.6	235	1 SMS	parotid secretory
2	336.5	27.3	235	2 B42337	parotid secretory
3	212.5	17.2	206	2 A42337	submandibular gland
4	110.5	9.0	1075	2 T52638	exportin 1 (xla)
5	110.5	9.0	1075	2 T51557	Exportin (XPO1) p
6	109	8.8	1769	2 S53378	probable membrane
7	108	8.8	825	2 B89944	hypothetical prote
8	106.5	8.6	473	2 S17448	probable ligand-bi
9	106	8.6	296	2 A49943	fructose-bisphosph
10	104.5	8.5	1727	2 T50073	myosin-like coiled
11	104	8.4	1616	2 G42424	cytadherence-acces
12	104	8.4	2710	2 A37052	toxin A - Clostrid
13	103.5	8.4	296	2 F90067	hypothetical prote
14	100.5	8.2	1056	1 G02157	kinesin-like spind
15	100	8.1	990	2 C82759	zinc proteinase XF
16	99.5	8.1	1524	2 S85533	surface layer prot
17	99	8.0	679	2 D84182	DNA ligase (NAD+)
18	99	8.0	1441	2 B86807	hypothetical prote
19	98.5	8.0	1615	1 WMTM8T	180K protein - tom
20	98	7.9	521	2 S34338	hypothetical prote
21	97.5	7.9	211	2 D84505	biliary glycoprote
22	97.5	7.9	328	2 A82087	hypothetical prote
23	97.5	7.9	414	2 T27045	phosphoserine phos
24	97	7.9	757	1 S64742	hypothetical prote
25	97	7.9	958	2 S64249	dynamitin-related pr
26	96.5	7.8	1345	2 H90975	hypothetical prote
27	96.5	7.8	2660	2 B85822	probable invasin 2
28	96	7.8	458	2 JG1509	biliary glycoprote
29	96	7.8	624	2 PC6003	surface membrane p

ALIGNMENTS

RESULT 1

SQMS

parotid secretory protein precursor - mouse

N:Alternate names: PSP

C:Species: Mus musculus (house mouse)

C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 22-Jun-1999

C:Accession: A23031; I53236

R:Madson, H.O.; Hjorth, J.P.

Nucleic Acids Res. 13, 1-13, 1985

A:Title: Molecular cloning of mouse PSP mRNA.

A:Reference number: A23031; MUID:85215456

A:Accession: A23031

A:Molecule type: mRNA

A:Residues: 1-235 <MAD>

A:CROSS-references: GB:X01697; NID:q53810; PIDN:CAA25846.1; PID:q758163

R:Poulsen, K.; Jakobsen, B.K.; Mikkelsen, B.M.; Harmark, K.; Nielsen, J.T.; Hjorth, J.

EMBO J. 5, 1891-1896, 1986

A:Title: Coordination of murine parotid secretory protein and salivary amylase expres

A:Reference number: I53236; MUID:87004556

A:Accession: I53236

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-87 <RES>

A:CROSS-references: GB:M26807; NID:q200556; PIDN:AAA40009.1; PID:q554264

C:Comment: PSP is the most abundant protein in the parotid gland. Its function is not

C:Genetics:

A:Gene: psp

A:Map position: 2

A:Introns: 41/1

A>Note: list of introns may be incomplete

C:Superfamily: parotid secretory protein

C:Keywords: parotid gland; saliva

F:1-20/Domain: signal sequence

F:21-235/Product: parotid secretory protein #status predicted <SIG>

F:21-235/Product: parotid secretory protein #status predicted <MAT>

Query Match 29.6%; Score 364.5; DB 1; Length 235;
Best Local Similarity 33.8%; Pred. No. 1.9e-19;
Matches 81; Conservative 62; Mismatches 80; Indels 17; Gaps 3;

QY 1 MLQWLKLVLLCGVLTGTSSESLLDNLGNLNNVVDKLEPVLHGLETVDTNTLKGLEKLKY 60

DB 1 MFQGLSVLVCGLLGNSSLSLGGELGSAVNN-----LKLNPPEAVPQNLNL 48

QY 61 DLGVLOKSSAWLQAKQAEKLLNNVLSKLLPTNTDIFG--LKISNLLLDVKAEPID 118

DB 49 DVLELQQAATSWPLAKNSIILET---LNTADLGNLKSFTSLNGLLLKLNKLKLVLDQAKLSS 105

QY 119 DGKGLNLSFPVTANTVAGPIGTQIINKASLDLFTAVTETDPTQHPQAVLGEACADP 178

DB 106 NNGINDLTVPLAGEASLVLPFGICKTVDISVSLDNLNSIKNTNAQTGLPEVTICKCSNT 165

hypothetical prote
cell proliferation
probable DNA topoi
methyl-accepting c
hypothetical prote
hypothetical prote
conserved hypotet
ribosomal RNA proc
nitrogenase (EC 1.
ketol-acid reducto
hypothetical prote
uncharacterized co
NIF80 protein - ye
hypothetical prote
biliary glycoprote
delta-l-pyrroline-

Query Match: 17.28; Score 212.5; DB 2; Length 206;

RESULT 5

T51557

Exportin1 (XP01) protein - Arabidopsis thaliana (fragment)
 N:Alternate names: protein F2K13_170
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 16-Feb-2001
 C:Accession: T51557
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
 submitted to the Protein Sequence Database, August 2000
 A:Reference number: T25394
 A:Accession: T51557
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1075 <SAV>
 A:Cross-references: EMBL:AL391141
 A:Experimental source: cultivar Columbia; BAC clone F2K13
 C:Genetics:
 A:Map position: 5
 A:Superfamily: Arabidopsis thaliana exportin 1

Query Match 9.0%; Score 110.5; DB 2; Length 1075;
 Best Local Similarity 23.8%; Pred. No. 4.5;
 Matches 68; Conservative 46; Mismatches 107; Indels 65; Gaps 15;

QY 6 KLVLGQ-----VLGTSESLLDNLGNVVDKLEP-VLHEGLETVDTNLKLGILEK 59
 DB 593 KIVOKARKKRVTVQNGEPFSELTGLTATTVODLEPHQIHFSYESVGNMIQAESDPQK 652
 QY 60 VDLGVQKSSA-----W-QLAKQKAGEAEKLLNNVSK-----LLPTNTD-----IFGLK 103
 DB 653 RD-EYLQRLMALPNQKWAELIGQARHSVEFLKQDVVIRVNLQNTNSAATSLGTFLS 711
 QY 104 INSNLLDNKAEPIDGKGLNL-----SFPTVNVVAGPIIQ-----IINKASLDLLTAVT 157
 DB 712 QLSLFLDM-----LNVRYMSYSELVSTNITEGGPYASKTSFVKLLRSVKRETCLK 761
 QY 158 IET--DPQTHQ-----VAVLGECASD-PTSISLSLLDKHSOTINKFVNSVINT 203
 DB 762 TETFLDKAEDPHGKQFVPPMMESVLGDYARNVPDAREVLSLFAFLINKYKATMLQD 821
 QY 204 LKSTVSSLLQKEICPLIRIFISLSDVNVVQVVDNPQHKTLQTLTI 249
 DB 822 VPHIFEAVFQ---CTLEMI-----TKNFEDYPEHRLKPFSL 855

RESULT 6

S53378
 probable membrane protein YJL109c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein J0808
 C:Species: Saccharomyces cerevisiae
 C:Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
 C:Accession: S53378; S56887; S57359
 R:Rasmussen, S.W.
 submitted to the EMBL Data Library, February 1995
 A:Description: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and
 A:Reference number: S53376
 A:Accession: S53378
 A:Molecule type: DNA
 A:Residues: 1-1769 <RAS>
 A:Cross-references: EMBL:X85021; NID:g728698; PID:g728701
 R:Rasmussen, S.W.
 submitted to the Protein Sequence Database, September 1995
 A:Reference number: S56876
 A:Accession: S56887
 A:Molecule type: DNA
 A:Residues: 1-1769 <RAW>
 A:Cross-references: EMBL:249384; NID:g1008293; PID:g1008293; MIPS:YJL109c
 R:Rasmussen, S.W.
 Yeast 11, 873-883, 1995
 A:Title: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and CSD3

A:Reference number: S57357; MUID:96090136

A:Accession: S57359
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1769 <RAF>
 A:Cross-references: EMBL:X85021; NID:g728698; PID:CAA59385.1; PID:g728701
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
 C:Genetics:
 A:Map position: 101
 C:Keywords: transmembrane protein

Query Match 8.8%; Score 109; DB 2; Length 1769;
 Best Local Similarity 19.2%; Pred. No. 11;
 Matches 52; Conservative 48; Mismatches 93; Indels 78; Gaps 9;

QY 20 SLLDNLGNLSNVVDKLEPVLHEGLETVDTNLKLGILEKLVGLCVLQKSSAWQLAKQAK 79
 DB 331 SILTFLDKEDKPCDKFTTSTRSIAHYDRS-----KLNILSLKK-----IRLRY 378
 QY 80 EAEXLLNNVJ--SKLLPTNTDI-----FGLKISNLSLIDVKAEPIDDGKGLNLSFPTV 132
 DB 379 EVRLITDILVLSLEDKXSQVLEFEYFISINEDLVKLK-----KSLGLT----- 425
 QY 133 VTVAGPIIQIINKASLDLLTAVTIED--PQTHQPAVLGECASDPTSISLSLDKHS 190
 DB 426 -----GELFEIRLTLSLTNADVNTDIVKQLSDPV-----ETTKKDTASFQTF 473
 QY 191 QIIN-----KFNVSIVNTLKSIVSSLLQKEICP 218
 DB 474 ELINTTVSMLETGERYKKVLSLFTETAGKGYKASSFLTSTFTLESITFLLRVTISP 533
 QY 219 LIRIFHSLDVNVVQVVDNPQHKTLQTLI 249
 DB 534 AAPALKLIISLNIKIYINSIEKVNIFTLV 564

RESULT 7

B89944
 hypothetical protein SA1447 [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: B89944
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 337, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: B89944
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-825 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701417; PIDN:BA642711.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA1447

Query Match 8.8%; Score 108; DB 2; Length 825;
 Best Local Similarity 24.7%; Pred. No. 4.8;
 Matches 72; Conservative 39; Mismatches 94; Indels 86; Gaps 15;

QY 17 TSSSLDNLGNL-----SNVVDKLEPVLHEGLETVDTNLKLGILEKLVGLQK 67
 DB 89 TKEAIISSLSLDFKGVGKKTQAINVTLG-----DNAINDILD-----DHSVLEK 134
 QY 68 SSWQLAKQKAGEAKL-LNNVYSKLLPTNTDI-FGLKISN----- 107
 DB 135 VSLGSKKKKQ-QIAEQISANSESEKIMRLHDLGFGPKLSMAIYQFVLGDTLTILDRNPY 193
 QY 108 -LILDVKAEPIDDGKGLNLSFPTVNVVAGPIIQIINKASLDLLTAVTIED----P 162

Db 194 QLIYDIK-----GICFNKADOLARNIGTA---YNDNERLKAAL-----LYTLEEECIKQG 240
 QY 163 QHPQPAVLGECA-----SDPTSLSLLDKHQSILNKFNVSINTLKSTVSSILKKEI 216
 Db 241 HTYLPINVIDTVLVNYQDEVEIPEKLBMLQYLNEERLLIDNEQVAIPSLYSEI 300
 QY 217 CP---LIRIFHSIDVNVVYQ-----VVDNPQ-----HKTOLOTLI 249
 Db 301 KSVQNLFRITKTHNKLTELEQSDLMQIHGTEIDANQVNYAASQREALQTAI 351

RESULT 8

S17448

probable ligand-binding protein RYA3 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999

C:Accession: S17448

R:Dear, T.N.; Boehm, T.; Keverne, E.B.; Rabbitts, T.H.

EMBO J. 10, 2813-2819, 1991

A:Title: Novel genes for potential ligand-binding proteins in subregions of the olfactory

A:Reference number: S17447; MUID:92007724

A:Accession: S17448

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-473 <DEA>

A:Cross-references: EMBL:X60658; NID:q57733; PIDN:CAA43065.1; PID:q57734

Query Match

Best Local Similarity 8.6%; Score 106.5; DB 2; Length 473;

Matches 60; Conservative 53; Mismatches 93; Indels 81; Gaps 12;

QY 1 MLQMKLVLLCGVLTGTSESLL----DNLGNDLSNVVDKLEPVLHGLEIVDNTLKGLE 56

Db 8 LLLMLGLATPCUGLLETGTGTARIDKDELGAIONSLVG-GPILQNLVGLVTSVNOGLLG 66

QY 57 K--LXVDLVGVLSKSAWQAKOKAQAEKLNNVISKLLPTNTDIFGLKISNLI--LDV 112

Db 67 AGGLLGGGGLLSYGLFLSVE-----ELSGLKIEELTLPVSI 104

QY 113 KAEPIDCKGKLNLSFPVNTVNTVAGPIIGIINKASLDLLTAVTIEDTQHPQPAVLG 172

Db 105 KLLP---GVGVLSHTKVSLSHSGPLVG-LQLAAEVNVSCKVALGMSPR-GTPILILK 159

QY 173 ECASDPTSLS-----LLDKHSQIINKFNVSIN---- 202

Db 160 RCNTLLGHISLTSGLLPTPTFGLVEQTLCKVLPGLLCPVVDLSVNVNELLGATLSLVL 219

QY 203 ----TLKSTVSSLLAKKEICPLIRIFHSIDVNVYQ-----VVDNPQ 240

Db 220 GPLGSVEPTLATL-----PLISNOYIELDINPIVKSIAQDVIDFPK 260

RESULT 9

A49943

fructose-bisphosphate aldolase (EC 4.1.2.13) - Staphylococcus carnosus (strain TM300)

C:Species: Staphylococcus carnosus

C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 15-Oct-1999

C:Accession: A49943; S33358

R:Witke, C.; Goetz, F.

J. Bacteriol. 175, 7495-7499, 1993

A:Title: Cloning, sequencing, and characterization of the gene encoding the class I fruc

A:Reference number: A49943; MUID:94042930

A:Accession: A49943

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-296 <WIT>

A:Cross-references: EMBL:X71729; NID:q297873; PIDN:CAA50663.1; PID:q297874

C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match

Best Local Similarity 8.6%; Score 106; DB 2; Length 296;

Matches 60; Conservative 53; Mismatches 93; Indels 81; Gaps 12;

Matches 50; Conservative 31; Mismatches 64; Indels 48; Gaps 9;
 QY 52 KQILEKLVLDGLGVLOKSSAWQAKAQAEKLNNVISKLLPTNTDIFGLKISNLIID 111
 Db 94 KGIPEKLVKDKGLAEEAGVQLMK-PIPDLDKLLDRA-----NERGIFGTM-KSNILE 145
 QY 112 VKAEPIDDDKGLNLSFPVNTVNTVAG--PIIGIIN-----LKASLIID-- 151
 Db 146 NNKEALE--KVVKQQFEVAKKEIIAAGLVPIITEPVNINAKOKEATEANLAKAELIINL 203
 QY 152 -----LLTAVTIET-----DPOTHQPAVLGECASDPTSISLSLLDKHSQIINKF 196
 Db 204 KKDQYVMLKLTITPKVNAVYSELIEHPQVIRVVALSGVSRDEAN---KILKQNDGLIASF 260
 QY 197 VNSVINTLKSTVS 209
 Db 261 SRALVSDLNAQOS 273

RESULT 10

T50073

myosin-like coiled-coil protein sp8 [imported] - fission yeast (Schizosaccharomyces p

C:Species: Schizosaccharomyces pombe

C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000

C:Accession: T50073

R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.

submitted to the EMBL Data Library, December 1999

A:Reference number: Z25034

A:Accession: T50073

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1727 <MCD>

A:Cross-references: EMBL:AL133357; PIDN:CAB62414.1; GSPDB:GN00066; SPDB:SPAC1486.04c

A:Experimental source: strain 972h(-); cosmid c1486

C:Genetics:

A:Gene: SPDB:SPAC1486.04c

A:Map position: 1

Query Match

Best Local Similarity 8.5%; Score 104.5; DB 2; Length 1727;

Matches 63; Conservative 41; Mismatches 94; Indels 61; Gaps 11;

QY 18 SFSLDLNGLNDSNVVDKLEPV-----LHEGLETVDN----- 49

Db 333 SENSRELEQEKYDSVSELOVVKENKNTSVSAGVGLFPLAQKLSAVONPEFSTKYVSD 392

QY 50 --FLAGILLEKLVLDGLVLOK--SSAWQLAKQ-----KAQAEKLNNVISKLLPTNTDIF 100

Db 393 NMKIQKVYSSILKQLDLRLTNKPSFCEQVKQKIPVVKQQRSEIVRNNIYMNFLSES---- 448

QY 101 GKISNSLIIDVKAEPIDDDKGLNLSF-PVTANVTVAGPIIGIINKASLDLLTAVTIE 159

Db 449 -LETNNNLTQVQALLSTKVRQEAACYLQLTASRTQCSDLRSREVICLMAELDHLNETKSR 507

QY 160 TDPQTHQPAVLGECASDPTSISLSLLDKHSQIINKFNVSINTLKSTVSSLLQKEICPL 219

Db 508 NVPATVQ--VALDEVAQNPSASTETLVNK--ELAN-----FSSIKEAVSKTLE----- 551

QY 220 IRIFIHSL--DVNVYIQVY 236

Db 552 LREKVRALCEDVDVETQKQTV 570

RESULT 11

G54242

cytochrome-accessory protein (hmw1) homolog MG386 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999

C:Accession: G54242

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.

M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,

C.A.; Venter, J.C.

Science **270**, 397-403, 1995
A:Title: The minimal gene complement of *Mycoplasma genitalium*.
A:Reference number: A64200; MUID:96026346
A:Accession: G64242
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1616 <TIGR>
A:Cross-references: GB:U39723; GB:U43967; NID:g1046092; PID:g1046097; TIGR:MG386
A:Experimental source: strain G-37
A:Genetics: C:Genetics
A:Genetic code: SGC3

Query Match 8.4%; Score 104; DB 2; Length 1616;
Best Local Similarity 20.6%; Pred. No. 23;
Matches 53.

```

14
07 ANGELS 120; GAPS
18 SESLLDNLGNDLSNVVDKLEPVLHEGLETVDNLTGKILEKLVGLGVOKSSAWOLAKQ 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 SKEIKDASKAULSNISDDIDSVNKFCGSPFDEKQSVKRSQVDEIILDANNDF--INE 517
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 AQAEAEKLNNVIS-----KLLPTNTD----- 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 SULPREVVNNIDSOINTVSEQFPTYSVNEFQCFSEPPVVSDEKIKETNSDESVENTDL 577
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 --IFGLKISNLSLDVKAEPIDDGKGLNLSFPVTANVTVA-----GPIGOI--IN--- 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 TALFEKLVNEVLL--TNEVVD---VNAPPSTETEVKVSSELPKSELVDEITFINNDPK 631
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 ---LKASLDLL-----TAVTITDTPQTHOPVAVL-----GECASDPTS 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 PQEGLEYKVDFLETEPKSLFDEKTTIIVSEPPFTIPLDLSLELDSVNDVKOKSLTETTSV 691
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 SLSLLDKHSQIINKFVNSVINTLKSTVSSLLQKEICPLRIFHSLDVNVITQGVVNQPH 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 692 ELN---HEEIGNEFIN-----LDYSE-KEVQSEOP--- 716
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 242 KTQLQT 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 717 TTOLET 722
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12
A37052 toxin A - Clostridium difficile
C.Species: Clostridium difficile
C.Date: 31-Jan-1992 #sequence:revision 31-Jan-1992 #text_change 24-Sep-1999
C.Accession: A37052; A60991; S21897; S22437; S08638
C.H.: Wang, S.Z.; Price, S.B.; Phelps, C.J.; Lyerly, D.M.; Wilkins, T.D.; Johnson
Infect. Immun. 58, 480-488, 1990
A.Title: Molecular characterization of the Clostridium difficile toxin A gene.
A.Reference number: A37052; MUID:90129405

A: Molecule type: DNA
A: Residues: 1-2710 <DOV>
A: Cross-references: GB:M30307; NID:g144925; PIDN:AAA23283.1; PFD:g144926
R:Wren, B.W.; Clayton, C.L.; Tabagchali, S.
FEMS Microbiol. Lett. 70, 1-6, 1990
A: Title: Nucleotide sequence of Clostridium difficile toxin A gene fragment and detection
A: Reference number: A60991
A: Accession: A60991
A: Molecule type: DNA
A: Residues: 1, 1894-1899, 'K', 1901-1910, 'Y', 1912-1919; 2054-2074; 2096-2116; 2138-2158 <WRE
A: Cross-references: GB:X17194
R: Eichel-Streiber, C.; Laufenberg-Teldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
Submitted to the EMBL Data Library, July 1991
A: Description: Comparative analysis of Clostridium difficile toxins A and B.
A: Reference number: S21894

Accession: S21897
A.Molecule type: DNA
A.Residues: 1-154 <EIC>
A.Cross-references: EMBL:X60984
R.von Elchei-Streiber, C.; Laufenberg-Feldmann, R.; Sarttingen, S.; Schulze, J.; Sauerbor

Mol. Gen. Genet. 233, 260-268, 1992

A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.

A:Reference number: S22434; MUID:92293124

A:Accession: S22437

A:Molecule type: DNA

A:Molecule type: DNA

A:Residues: 1-92 <VON>

A:Cross-references: EMBL:X60984

A>Note: The four fragments shown in reference A50991 correspond to four types of re d with repeats ordered ASCCCDABACBCCCDABCCDABCCDABCB

R:Sauerborn, M.; von Elchei-Streiber, C.

Nucleic Acids Res. 18, 1629-1630, 1990

A:Title: Nucleotide sequence of Clostridium difficile toxin A.

A:Reference number: S08637; MUID:90221894

A:Accession: S08638

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-55,'v', 57-2079,'l', 2081-2549,'S', 2551-2710 <SAU>

A:Cross-references: EMBL:X51757

C:Genetics:

A:Gene: toxA

C:Superfamily: Clostridium difficile toxin A; cpl repeat homology

C:Keywords: cytotoxin; enterotoxin

F:1830-1839/Domain: cpl repeat homology <CP01>

F:1840-1860/Domain: cpl repeat homology <CP02>

F:1861-1881/Domain: cpl repeat homology <CP03>

F:1933-1952/Domain: cpl repeat homology <CP04>

F:1953-1973/Domain: cpl repeat homology <CP05>

F:1974-1994/Domain: cpl repeat homology <CP06>

F:1995-2015/Domain: cpl repeat homology <CP07>

F:2067-2086/Domain: cpl repeat homology <CP08>

F:2087-2107/Domain: cpl repeat homology <CP09>

F:2108-2128/Domain: cpl repeat homology <CP10>

F:2129-2149/Domain: cpl repeat homology <CP11>

F:2201-2220/Domain: cpl repeat homology <CP12>

F:2221-2241/Domain: cpl repeat homology <CP13>

F:2242-2261/Domain: cpl repeat homology <CP14>

F:2315-2334/Domain: cpl repeat homology <CP15>

F:2335-2355/Domain: cpl repeat homology <CP16>

F:2356-2376/Domain: cpl repeat homology <CP17>

F:2377-2397/Domain: cpl repeat homology <CP18>

F:2449-2468/Domain: cpl repeat homology <CP19>

F:2469-2489/Domain: cpl repeat homology <CP20>

F:2490-2510/Domain: cpl repeat homology <CP21>

F:2562-2581/Domain: cpl repeat homology <CP22>

F:2582-2602/Domain: cpl repeat homology <CP23>

F:2653-2672/Domain: cpl repeat homology <CP24>

F:2673-2694/Domain: cpl repeat homology <CP25>

Query Match: 8.4%; Score 104; DB 2; Length 2710;
Best Local Similarity 20.3%; Pred. No. 45;
Matches 56; Concentration 1.00

```

QY 6 KVLVLCGVLTGTSLSLDNIQNDLSNVV-----DKLEPVHCELETVDNLTG--GILEK 57
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 807 KTLILDASVSDPTKFIINLKLNIETSSIGDYIYEKLEPVKNIHNSIDDLIDENLLEN 866
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 58 LKVDLGVLQK-----SSAQLAKOQAEAKLLN--NVISKLLPTNTDIFG----- 101
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 867 VSEDEYELKLNLLNDEKYLISFEDISKNNSTYSVRFINKNGESVYVETEKEIFSKYSEH 926
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 102 -----LKISNLSLIDVKAEPIDGKGLNLSFPVTANTVAGPIGQIINLKSALDLLTAV 156
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 927 ITRKISTIKNSIITDNGNLLDN---IQDHTSQVNTLNAAFFIQSLIDYSSNKVDLNDL 983
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 157 TITDPTQHPVAVLGECASDPTSISSLDDKHSQIINKFVNSVINTLKTSTVSSLLQKEI 216
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 984 STSVKVKOLH---AQLFSTGLNTIYDSIQLVN---LISNAVNDTINVLPITTEGI----- 1031
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 217 CPLIRIFHSLDWN-VIOQVVD--NPQHKTLQTLI 249
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1032 -PIVSTILDGICNLGRAINELLDEHPDLKLKEARV 1066
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 13
 F90067
 hypothetical protein SA2399 [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: F90067
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: AB9756; MUID:21311952; PMID:11418146
 A:Accession: F90067
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-296 <KUR>
 A:Cross-references: GB:BA000018; PID:gl3702563; PIDN:BA043704.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA2399

Query Match 8.4%; Score 103.5; DB 2; Length 296;
 Best Local Similarity 25.5%; Pred. No. 2.7;
 Matches 53; Conservative 30; Mismatches 68; Indels 57; Gaps 10;

QY 42 EGLTEVDNLT-KGIIKLIKVDLGVLOKSSAWOLAKQAKQAEKLLNNVSKLLPTN-TDI 99
 DB 83 EKKYADYLDAGGVFFKLVKDGAEQNGVLMKP-----IDNLSLLDRANRHI 134
 QY 100 FGLKISNLSILDVKAEPIDDDKGLNLSFPVTANVTAG--PIIGIIN----- 145
 DB 135 FQTKM-RSNILELNQGIKD--VVEQQEVAQIIAKGLVPIIEPVNINAKDKAEIEKV 191
 QY 146 -----LKASLDLTA-----VTTETDQTHQPVAVLGEASDPTSL----- 183
 DB 192 LKAEILKGLGSLNADQLVMLKLTPTPEMLYK-----ELAEHPNVVVRVVLGGYSREK 245
 QY 184 ---SLLDKHSQIINKFVNSVINTLKSTS 209
 DB 246 ANELKNDDELTAJSFSLALSDLRADQS 273

RESULT 14
 G02157
 kinesin-like spindle protein HKSP - human
 C:Species: Homo sapiens (man)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: G02157
 R:Whitehead, C.
 Submitted to the EMBL Data Library, September 1995
 A:Reference number: H00839
 A:Accession: G02157
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1056 <WHI>
 A:Cross-references: EMBL:U37426; NID:gl171152; PIDN:AAA86132.1; PID:gl171153
 C:Genetics:
 A:Gene: GDB:KNSL1; E95; KSP
 A:Cross-references: GDB:132856; OMIM:148760
 A:Map position: 10q24.1-10q24.1
 C:Superfamily: kinesin-related protein eg5; kinesin motor domain homology
 C:Keywords: ATP; nucleotide binding; P-loop
 F:19-365/Domain: kinesin motor domain homology <KMT>
 F:105-112/Region: nucleotide-binding motif A (P-loop)

Query Match 8.2%; Score 100.5; DB 1; Length 1056;
 Best Local Similarity 23.0%; Pred. No. 24;
 Matches 65; Conservative 39; Mismatches 89; Indels 89; Gaps 14;

QY 22 LDNLQNLNSNVVDKLEPVLHEGLETVNTLKGILEKLVLDGLVQLK--SSAWOLAKQKRAQ 79

DB 447 LDCKSDLQNKTOELE-----TTQKHLQETKQL--VKREYITSALEETEEKLH 493
 QY 80 E-AEKLNNVSKLLPTNTDIFGL--KISNSULLDVKAEPIDD--GKGLNLSF----- 127
 DB 494 DAASKLLNTVEE-----TTKDVSGLSKLOKKKAVDOHNAEAQDIFGKNLSLNNMKEFL 549
 QY 128 ----PVTANVTAGPIIGIINLKAS-LDLLTAVTITETDQTHQPVAVLGEASDPTSL 181
 DB 550 KDSSQKQKAMCEVHKTLFGNLLSSVSALDTITV-----ALGSIITSIHENV 596
 QY 182 SLSLLDKHSQIINKF-----VNSVINTLKSTVSSLIQKEICPLI----- 220
 DB 597 STHV-----SQIFNMILKEOSLAESKTVLQELINVLKLTCLLSLEMLSPVVSILKINS 652
 QY 221 ---RIFIHSL-----DVNVIOQVVDNPKHKTQLOTL 248
 DB 653 QLKHIPTSLTVADKTEDOKKELDGFLSILCNLNLHELQENTI 694

RESULT 15
 C82759
 Zinc proteinase XF0816 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82759
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: AB2515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: C82759
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-990 <SIM>
 A:Cross-references: GB:AB003921; GB:AE003849; NID:99105710; PIDN:AAF83626.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrei
 as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
 Submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; France, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; L
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF0816

Query Match 8.1%; Score 100; DB 2; Length 990;
 Best Local Similarity 21.3%; Pred. No. 24;
 Matches 54; Conservative 45; Mismatches 105; Indels 50; Gaps 8;

QY 15 TCTSESLDNLGNDLSNVVDK-LEPVLHEGLETVNTLKGILEKLVLDGLVLOKSSAWQL 73
 DB 706 SGTNEATQSLTATDLKDFQQRWLRPDNVRILVTGHTTLKSLIIPQLEAAFGDQWQAPSTKS 765
 QY 74 AKQAKAEKLLNVSKLLPTNTDIFGLKISNLSL-DVKAEPIDDKGLNLSFPVTAN 132
 DB 766 HKQ-----ITDVAQFKPRIFLIHRPEAQOQLLAGLAPTKD-----PANLE 809
 QY 133 VTVAGPIIGIINLKASLDL-----LTAVTETDQTHQPVAVLGEASDPTSLSL 185
 DB 810 INVNEAFGCTPSSRLNMLREKHWAYGASSVLPNAQGORPVVFIAPVQTDKTAESIAE 869
 QY 186 LDKHSQ--IINKFVNSVINTLKSTVSSLLQKEICPLIRIFHSLD-----VNVVIOQ 234

Db 873 IQEAGQVIVNK-----PLTQEEVDKIKQIIRSLPGSYETSGAVLDAYES 915
QY 235 VVDNPOHKTQLOTL 248
Db 916 IVRYERPDNYIOTL 929

Search completed: August 6, 2002, 17:05:10
Job time: 912 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 17:26:36 : Search time 1822.42 seconds
(without alignments)
8566.190 Million cell updates/sec

Title: US-10-020-139-1_COPY_48_793

Perfect score: 746

Sequence: 1 GATCGTCAGCTTTGGAAAC.....AACCCAGCTCCAAACCTCA 746

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 250 summaries

Database :

GenEmbl:

1:	gb-ba:	1	746	100.0	1058	6	AX061621	Sequence
2:	gb-ba:	2	694	93.0	750	9	AF432917	Homo sapi
3:	gb-in:	3	159	21.3	161593	9	HSBA49G10	Human DNA
4:	gb-in:	4	23	3.1	999	4	BTU79413	Bos taurus
5:	gb-in:	5	23	3.1	1024	4	BTU79414	Bos taurus
6:	gb-in:	6	21	2.8	1730	9	AK024183	Homo sapi
7:	gb-in:	7	21	2.8	79043	9	HSBJ9614	Homo sapi
8:	gb-in:	8	21	2.8	95289	2	AC084416	Continuation (6 of
9:	gb-in:	9	21	2.8	100259	9	AC005800	Homo sapi
10:	gb-in:	10	21	2.8	110000	2	AC084416	Continuation (5 of
11:	gb-in:	11	21	2.8	157248	9	AC002312	Human Chr
12:	gb-in:	12	21	2.8	158769	9	AC094795	Rattus no
13:	gb-in:	13	21	2.8	166141	9	AL137059	Human DNA
14:	gb-in:	14	21	2.8	169468	2	AC097610	Rattus no
15:	gb-in:	15	21	2.8	169824	2	AP001854	Homo sapi
16:	gb-in:	16	21	2.8	171180	2	AC067986	Homo sapi
17:	gb-in:	17	21	2.8	172628	2	AC087675	Homo sapi
18:	gb-in:	18	21	2.8	183553	2	AC106518	Rattus no
19:	gb-in:	19	21	2.8	185983	2	AC098697	Rattus no
20:	gb-in:	20	21	2.8	188741	2	AC013459	Homo sapi
21:	gb-in:	21	21	2.8	199254	2	AC021144	Homo sapi
22:	gb-in:	22	21	2.8	202032	2	AC087399	Homo sapi
23:	gb-in:	23	20	2.7	4692	4	SSP101	scrofa mr
24:	gb-in:	24	20	2.7	4892	6	AR026672	Sequence
25:	gb-in:	25	20	2.7	4892	6	AR026678	Sequence
26:	gb-in:	26	20	2.7	4892	6	AR029061	Sequence
27:	gb-in:	27	20	2.7	4692	6	AR033965	Sequence
28:	gb-in:	28	20	2.7	36777	9	AL607106	Human DNA
29:	gb-in:	29	20	2.7	97192	2	AC107373	Homo sapi
30:	gb-in:	30	20	2.7	102781	2	AF176680	Homo sapi
31:	gb-in:	31	20	2.7	114599	2	AC095654	Rattus no
32:	gb-in:	32	20	2.7	134882	9	AC008622	Homo sapi
33:	gb-in:	33	20	2.7	144536	9	AL161454	Human DNA
34:	gb-in:	34	20	2.7	158276	2	AC027709	Homo sapi
35:	gb-in:	35	20	2.7	175335	2	AC094124	Rattus no
36:	gb-in:	36	20	2.7	176601	9	AC073581	Homo sapi
37:	gb-in:	37	20	2.7	190390	10	AC083892	Mus muscu
38:	gb-in:	38	20	2.7	208964	2	AL645842	Mus muscu
39:	gb-in:	39	20	2.7	210995	2	AL012250	Homo sapi
40:	gb-in:	40	20	2.7	216038	2	AL662927	Homo sapi
41:	gb-in:	41	20	2.7	230372	2	AC073693	Mus muscu
42:	gb-in:	42	20	2.7	240842	2	AC079485	Mus muscu
43:	gb-in:	43	19	2.5	788	14	S81230	UL52' helic
44:	gb-in:	44	19	2.5	2582	1	NGPRIOPA	N gonorrhoe
45:	gb-in:	45	19	2.5	2957	6	BD008630	Compositi
46:	gb-in:	46	19	2.5	15389	1	BPLPSBLOC	X90711 Bordetella
47:	gb-in:	47	19	2.5	30425	9	AL670463	Human DNA
48:	gb-in:	48	19	2.5	36339	9	HSB3387	S-pombe c
49:	gb-in:	49	19	2.5	36484	8	SPAPBIA10	Streptococ
50:	gb-in:	50	19	2.5	41642	1	BBH007747	Streptococ
51:	gb-in:	51	19	2.5	43075	7	UH8974	Streptococ
52:	gb-in:	52	19	2.5	58747	2	AC102174	Mus muscu
53:	gb-in:	53	19	2.5	59314	9	AL392047	Human DNA
54:	gb-in:	54	19	2.5	62832	2	AC107830	Mus muscu
55:	gb-in:	55	19	2.5	73648	2	AC109187	Mus muscu
56:	gb-in:	56	19	2.5	95760	9	AL590432	Human DNA
57:	gb-in:	57	19	2.5	106253	9	AP003815	Oryza sat
58:	gb-in:	58	19	2.5	106763	9	AP002091	Homo sapi
59:	gb-in:	59	19	2.5	125018	2	AC094856	Rattus no
60:	gb-in:	60	19	2.5	129641	9	AC008560	Homo sapi
61:	gb-in:	61	19	2.5	129856	9	AC016620	Homo sapi
62:	gb-in:	62	19	2.5	138725	2	AC012031	Homo sapi
63:	gb-in:	63	19	2.5	139629	2	AP003809	Oryza sat
64:	gb-in:	64	19	2.5	139848	2	AP004335	Oryza sat
65:	gb-in:	65	19	2.5	142388	2	AC011130	Homo sapi
66:	gb-in:	66	19	2.5	142603	2	AC103905	Canis fam
67:	gb-in:	67	19	2.5	145899	2	AC016611	Homo sapi
68:	gb-in:	68	19	2.5	151548	2	AC106459	Rattus no
69:	gb-in:	69	19	2.5	154984	9	AC019246	Homo sapi
70:	gb-in:	70	19	2.5	156784	2	AC016390	Homo sapi
71:	gb-in:	71	19	2.5	157959	9	AP003101	Homo sapi
72:	gb-in:	72	19	2.5	160355	2	AC022644	Homo sapi
73:	gb-in:	73	19	2.5	160725	2	HS919B11	Homo sapi

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	746	100.0	1058	AX061621	Sequence

C 74	19	2.5	160903	9	AC092812	Homo sapi	C 147	18	2.4	79874	2	AC027252	AC027252 Homo sapi
C 75	19	2.5	160990	9	AL157702	Human DNA	C 148	18	2.4	80568	9	AC002420	AC002420 Homo sapi
C 76	19	2.5	161857	9	AC008828	Homo sapi	C 149	18	2.4	82279	9	AC010464	AC010464 Homo sapi
C 77	19	2.5	162280	2	AC110617	Homo sapi	C 150	18	2.4	83279	9	AC079858	AC079858 Homo sapi
C 78	19	2.5	162390	2	AC103901	Canis fam	C 151	18	2.4	83279	9	AL445223	AL445223 Human DNA
C 79	19	2.5	162839	9	AL137846	Human DNA	C 152	18	2.4	88840	2	AC107377	AC107377 Homo sapi
C 80	19	2.5	163456	9	AC017071	Homo sapi	C 153	18	2.4	89000	9	AC016932	AC016932 Homo sapi
C 81	19	2.5	167541	2	AC104244	Homo sapi	C 154	18	2.4	91885	9	AC006055	AC006055 Homo sapi
C 82	19	2.5	167679	9	AC007570	Homo sapi	C 155	18	2.4	92141	2	AC021565	AC021565 Homo sapi
C 83	19	2.5	170969	9	AC027663	Homo sapi	C 156	18	2.4	92141	2	AC021565	AC021565 Homo sapi
C 84	19	2.5	171095	2	AC104008	Homo sapi	C 157	18	2.4	92407	6	AX067461	AX067461 Sequence
C 85	19	2.5	172161	8	AP003301	Oryza sat	C 158	18	2.4	94341	9	AL158036	AL158036 Human DNA
C 86	19	2.5	174162	2	AC027409	Homo sapi	C 159	18	2.4	97876	8	APT25P22	APT25P22 Arabidops
C 87	19	2.5	182970	2	AC069234	Homo sapi	C 160	18	2.4	105574	2	AC013308	AC013308 Homo sapi
C 88	19	2.5	183447	2	AC095856	Homo sapi	C 161	18	2.4	106795	2	AC095833	AC095833 Rattus no
C 89	19	2.5	186388	9	AP001189	AP001189 Homo sapi	C 162	18	2.4	110000	2	CEY105E8_4	CEY105E8_4 Mus muscu
C 90	19	2.5	192277	2	AC094345	Rattus no	C 163	18	2.4	110000	2	CEY105E8_4	CEY105E8_4 Mus muscu
C 91	19	2.5	193552	2	AC021569	Homo sapi	C 164	18	2.4	110000	2	LMFLCHR34_05	LMFLCHR34_05 Continuation (5 of
C 92	19	2.5	195931	2	AC068174	Homo sapi	C 165	18	2.4	110000	2	LMFLCHR34_06	LMFLCHR34_06 Continuation (7 of
C 93	19	2.5	195931	2	AC068174	Homo sapi	C 166	18	2.4	110000	2	LMFLCHR36_33	LMFLCHR36_33 Continuation (34 of
C 94	19	2.5	196540	2	AC025408	Homo sapi	C 167	18	2.4	110939	9	CNSO1DSW	CNSO1DSW Human chr
C 95	19	2.5	196879	2	AC098722	Homo sapi	C 168	18	2.4	112292	9	HSJ23BD15	HSJ23BD15 Human DNA
C 96	19	2.5	197843	2	AC024373	Homo sapi	C 169	18	2.4	113271	9	AL592466	AL592466 Human DNA
C 97	19	2.5	198364	9	AC092601	Homo sapi	C 170	18	2.4	114192	2	AC096010	AC096010 Rattus no
C 98	19	2.5	203050	2	HS4AN10	297197 Homo sapien	C 171	18	2.4	115348	2	AL357042	AL357042 Human DNA
C 99	19	2.5	204938	2	AC097034	AC097034 Rattus no	C 172	18	2.4	115348	2	AC08053	AC08053 Homo sapi
C 100	19	2.5	220495	9	AP000886	AP000886 Homo sapi	C 173	18	2.4	116501	2	AC099851	AC099851 Homo sapi
C 101	19	2.5	226785	2	AC079638	Homo sapi	C 174	18	2.4	117193	2	AC099851	AC099851 Homo sapi
C 102	19	2.5	252840	2	AL590864	Mus muscu	C 175	18	2.4	117641	2	AC094035	AC094035 Rattus no
C 103	19	2.5	258240	2	AL662812	Mus muscu	C 176	18	2.4	118445	2	AC096954	AC096954 Rattus no
C 104	19	2.5	332635	1	AP003005	AP003005 Mesorhizo	C 177	18	2.4	121656	9	AL357054	AL357054 Human DNA
C 105	18	2.4	366820	1	AP003008	AP003008 Mesorhizo	C 178	18	2.4	122969	2	AC105029	AC105029 Homo sapi
C 106	18	2.4	409	33	AC087005	AC087005 Giardia i	C 179	18	2.4	123943	9	AC006208	AC006208 Homo sapi
C 107	18	2.4	471	1	EAJ10010	AJ010010 Enterobac	C 180	18	2.4	129842	9	HS151B14	HS151B14 Human DNA s
C 108	18	2.4	720	3	DMU05982	U05982 Drosophila	C 181	18	2.4	130077	9	AL353643	AL353643 Human DNA
C 109	18	2.4	814	33	CNS01CJ5	AL115593 Botrytis	C 182	18	2.4	132693	2	AC106565	AC106565 Rattus no
C 110	18	2.4	852	1	ECJ227966	ECJ227966 Enterobac	C 183	18	2.4	132693	2	AC087367	AC087367 Homo sapi
C 111	18	2.4	852	1	ECJ227966	ECJ227966 Enterobac	C 184	18	2.4	133139	9	AC084853	AC084853 Homo sapi
C 112	18	2.4	858	1	ECJ227967	ECJ227967 Enterobac	C 185	18	2.4	134362	2	AC084853	AC084853 Homo sapi
C 113	18	2.4	858	1	ECJ227967	ECJ227967 Enterobac	C 186	18	2.4	134580	9	AC002526	AC002526 Human PAC
C 114	18	2.4	1001	11	G66686	G66686 Xq4042 KWOK	C 187	18	2.4	134781	10	AL355177	AL355177 Mouse DNA
C 115	18	2.4	1023	33	AC050043	AC050043 Giardia i	C 188	18	2.4	135224	2	AC022373	AC022373 Homo sapi
C 116	18	2.4	1444	1	MPU22331	U22331 Metabacteri	C 189	18	2.4	135958	2	AC078904	AC078904 Homo sapi
C 117	18	2.4	2015	3	DMOSBP1	U81502 Drosophila	C 190	18	2.4	136541	2	AP000485	AP000485 Homo sapi
C 118	18	2.4	2345	6	AR152005	AR152005 Sequence	C 191	18	2.4	137468	8	AP003269	AP003269 Oryza sat
C 119	18	2.4	2345	6	E33186	E33186 pIGR-1 belo	C 192	18	2.4	137733	2	AC026292	AC026292 Homo sapi
C 120	18	2.4	2458	9	BC015581	BC015581 Homo sapi	C 193	18	2.4	138021	9	AC090820	AC090820 Homo sapi
C 121	18	2.4	2784	10	KNFX111A	Y12502 R.norvegicu	C 194	18	2.4	138953	9	AC022778	AC022778 Homo sapi
C 122	18	2.4	2948	1	ECJ22736	AF144731 Enterobac	C 195	18	2.4	139431	2	AC084003	AC084003 Homo sapi
C 123	18	2.4	2968	10	AF144731	AF144731 Rattus no	C 196	18	2.4	141172	9	AL512600	AL512600 Human DNA
C 124	18	2.4	3206	10	D78303	D78303 Rattus norv	C 197	18	2.4	142446	9	AL512600	AL512600 Human DNA
C 125	18	2.4	3224	6	E13890	E13890 Rat Y521 g	C 198	18	2.4	142446	9	AL512600	AL512600 Human DNA
C 126	18	2.4	3537	3	AF195494	AF195494 Drosophila	C 199	18	2.4	142707	9	HSJ279A18	HSJ279A18 Human DNA
C 127	18	2.4	4321	3	AF145682	AF145682 Drosophila	C 200	18	2.4	144184	2	AL138732	AL138732 Homo sapi
C 128	18	2.4	4372	9	AK024375	AK024375 Homo sapi	C 201	18	2.4	146150	2	AC068333	AC068333 Homo sapi
C 129	18	2.4	4372	9	AK024375	AK024375 Homo sapi	C 202	18	2.4	147405	9	AP001993	AP001993 Homo sapi
C 130	18	2.4	5894	3	ACMHC	Y06624 Acanthamoeb	C 203	18	2.4	147834	2	AC106809	AC106809 Homo sapi
C 131	18	2.4	7419	1	VPTPOP	X17149 V. parahaem	C 204	18	2.4	147911	2	AC096947	AC096947 Homo sapi
C 132	18	2.4	7555	14	APCCG	M58152 Apple chlor	C 205	18	2.4	148883	9	HS358H7	HS358H7 Human DNA s
C 133	18	2.4	10192	1	U32769	U32769 Haemophilus	C 206	18	2.4	150230	3	CEY105E8E	CEY105E8E Human DNA
C 134	18	2.4	12917	9	AL662851	AL662851 Human DNA	C 207	18	2.4	150741	3	CEY105E8E	CEY105E8E Human DNA
C 135	18	2.4	15596	1	AE006987	AE006987 Mycobacte	C 208	18	2.4	151152	2	AC020656	AC020656 Homo sapi
C 136	18	2.4	25143	2	AC108575	AC108575 Rattus no	C 209	18	2.4	152533	2	AC016506	AC016506 Homo sapi
C 137	18	2.4	38970	1	MTCH1002	Z92539 Mycobacteri	C 210	18	2.4	152581	9	AC093900	AC093900 Homo sapi
C 138	18	2.4	39549	2	AC105140	AC105140 Homo sapi	C 211	18	2.4	152855	9	AC027671	AC027671 Homo sapi
C 139	18	2.4	48003	3	CEC11E4	Z81015 Caenorhabdi	C 212	18	2.4	154160	2	AC046154	AC046154 Homo sapi
C 140	18	2.4	51914	2	AC068022	AC068022 Homo sapi	C 213	18	2.4	155124	2	AL356112	AL356112 Homo sapi
C 141	18	2.4	58866	2	AC090594	AC090594 Homo sapi	C 214	18	2.4	155518	2	AC095374	AC095374 Rattus no
C 142	18	2.4	65126	2	AC102443	AC102443 Homo sapi	C 215	18	2.4	156277	9	AC026795	AC026795 Homo sapi
C 143	18	2.4	66015	2	AC009339	AC009339 Drosophila	C 216	18	2.4	158382	2	AC016507	AC016507 Homo sapi
C 144	18	2.4	66421	2	AC019757	AC019757 Drosophila	C 217	18	2.4	160197	2	OSJN00116	OSJN00116 Oryza sat
C 145	18	2.4	69479	9	AC008597	AC008597 Homo sapi	C 218	18	2.4	160759	2	AC094044	AC094044 Rattus no
C 146	18	2.4	71287	9	AL360180	AL360180 Human DNA	C 219	18	2.4	161863	9	AL606970	AL606970 Human DNA
			77233	9	AL360180	AL360180 Human DNA				162837	9	AL451137	AL451137 Human DNA

```

220 18 2.4 162949 2 AC068029 Homo sapi
c 221 18 2.4 163243 9 AL513304 Human DNA
222 18 2.4 163996 2 AC053501 Homo sapi
223 18 2.4 164916 9 AC011224 Homo sapi
224 18 2.4 165176 9 AC078941 Homo sapi
c 225 18 2.4 165299 9 HS5915 Homo sapi
c 226 18 2.4 167336 2 AC021350 Homo sapi
c 227 18 2.4 167663 9 AC093854 Homo sapi
c 228 18 2.4 167849 9 AC067815 Homo sapi
c 229 18 2.4 167939 9 AC008014 Homo sapi
c 230 18 2.4 167962 2 AL596063 Homo sapi
c 231 18 2.4 168373 9 AL391683 Homo sapi
c 232 18 2.4 168535 2 AC007172 Homo sapi
c 233 18 2.4 168616 2 AC013636 Homo sapi
c 234 18 2.4 168993 2 AC011584 Homo sapi
c 235 18 2.4 170227 2 AC019161 Homo sapi
c 236 18 2.4 170559 2 AC026989 Homo sapi
c 237 18 2.4 170707 2 AL589986 Homo sapi
c 238 18 2.4 171899 2 AC020562 Homo sapi
c 239 18 2.4 172115 2 AC099478 Homo sapi
c 240 18 2.4 172883 9 AC053466 Homo sapi
c 241 18 2.4 172963 2 AP003176 Homo sapi
c 242 18 2.4 173028 9 AC090371 Homo sapi
c 243 18 2.4 173534 9 AC007920 Homo sapi
c 244 18 2.4 173722 2 AL450993 Homo sapi
c 245 18 2.4 174128 2 AP001123 Homo sapi
c 246 18 2.4 174189 9 AC068185 Homo sapi
c 247 18 2.4 174334 3 AC011086 Drosophila
c 248 18 2.4 174923 2 AC023350 Homo sapi
c 249 18 2.4 175071 2 AC010274 Homo sapi
c 250 18 2.4 175202 2 AC027254 Homo sapi

```

ALIGNMENTS

```

RESULT 1
AX061621 1058 bp DNA linear PAT 24-JAN-2001
LOCUS
DEFINITION Sequence 26 from Patent WO0100806.
ACCESSION AX061621
VERSION AX061621.1 GI:12406703
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE dumas milne Edwards, J.B., Bougueleret, L. and Jobert, S.
JOURNAL Complementary dna's encoding proteins with signal peptides
Patent: WO 0100806-A 26 04-JAN-2001;
GENSET (FR)
FEATURES
source Location/Qualifiers
1..1058
/organism="Homo sapiens"
/db_xref="taxon:9606"
67..516
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC24982.1"
/db_xref="GI:12406704"
/translation="MLQWLKVLGVLGTSLLDNLGNDLSNVVDPKLEPVLHLEGL
ETVDNLKGTLEKLVKDLGVLOKSSAQLKAKAQAEKLLNNVLSKLLPNTDIFGL
KINSILDVKASPIODGKGLNLSFPVTANVTAGPIIGOIINLKASLDLITAVTET
DPDTHQFVAVLGECASDPISILSLDLKHSIINKFVNSVINTLSTVSSLLQKEICP
LIRIFTHSLDNLVNIQVVDNPQHKTLQTLI"
67..111
/note="Von Heijne matrix score 5.20 seq QLWLKVLVLCGLT/GT"
polya_signal 1023..1028
polya_site 1042..1058
BASE COUNT 291 a 302 c 231 g 234 t
ORIGIN

```

```

Query Match 100.0%; Score 746; DB 6; Length 1058;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGCTTCAGCTTTGGAAACCTTGCTCTCTGCGGGGTGCTCACTGGACCTCAGATC 60
Db 66 GATGCTTCAGCTTTGGAAACCTTGCTCTCTGCGGGGTGCTCACTGGACCTCAGATC 125
QY 61 TCTTCTTCACAACTCTTGCAATGACCTAAAGCAATCTCGTGATAGCTGGAACTGTCT 120
Db 126 TCTTCTTCACAACTCTTGCAATGACCTAAAGCAATCTCGTGATAGCTGGAACTGTCT 185
QY 121 TCACGAGGACTTGAGACAGTTGACAATCTCTTAAAGGATCTCTTGAAGAACTGAAGT 180
Db 186 TCACGAGGACTTGAGACAGTTGACAATCTCTTAAAGGATCTCTTGAAGAACTGAAGT 245
QY 181 CGACCTAGAGTGTTCAGAAATCCAGTCTTGCCAACTGCCCAAGCAGAGGCCAGGA 240
Db 246 CGACCTAGAGTGTTCAGAAATCCAGTCTTGCCAACTGCCCAAGCAGAGGCCAGGA 305
QY 241 AGCTGAGAAATGCTTGAACAATGTCATTTCTAAGCTGCTTCCAACTAACACGGACATTTT 300
Db 306 AGCTGAGAAATGCTTGAACAATGTCATTTCTAAGCTGCTTCCAACTAACACGGACATTTT 365
QY 301 TGGGTTGAAATCAGCACTCCCTCATCTCTGGATGTCAAAGCTGAACCGATCGATGATG 360
Db 366 TGGGTTGAAATCAGCACTCCCTCATCTCTGGATGTCAAAGCTGAACCGATCGATGATG 425
QY 361 CAAAGGCTTAACTGAGCTTCCCTGTGTCACGCGCAATGTCACTGTGCGGGGCCATCAT 420
Db 426 CAAAGGCTTAACTGAGCTTCCCTGTGTCACGCGCAATGTCACTGTGCGGGGCCATCAT 485
QY 421 TGGCCAGATTTCAACCTGAAAGCCTCTCTTGACCTCTGACCGCAGTCAACAATTGAAC 480
Db 486 TGGCCAGATTTCAACCTGAAAGCCTCTCTTGACCTCTGACCGCAGTCAACAATTGAAC 545
QY 481 TGATCCCGCAGACACACAGCCTGTTGCGCTCTGGGAGATGCGCCAGTCAACCAACCG 540
Db 546 TGATCCCGCAGACACACAGCCTGTTGCGCTCTGGGAGATGCGCCAGTCAACCAACCG 605
QY 541 CATCTCACTTTCTTGTGGACAAACACAGCAATCAATCAACAAGTTCGTGAATAGCGT 600
Db 606 CATCTCACTTTCTTGTGGACAAACACAGCAATCAATCAACAAGTTCGTGAATAGCGT 665
QY 601 GATCAACAGCTGAAACCACTGTATCTCTCTCTGCTCAGAGGAGATATGTCACCTGAT 660
Db 656 GATCAACAGCTGAAACCACTGTATCTCTCTCTGCTCAGAGGAGATATGTCACCTGAT 725
QY 661 CGGCATCTTCACTCACTCCCTGGATGCAATGTCATTCAGCAGGCTCGTCAATCTCTCA 720
Db 726 CGGCATCTTCACTCACTCCCTGGATGCAATGTCATTCAGCAGGCTCGTCAATCTCTCA 785
QY 721 GCACAAAACCCAGCTGCAAAACCTCA 746
Db 786 GCACAAAACCCAGCTGCAAAACCTCA 811

```

```

RESULT 2
AF432917 AF432917 750 bp mRNA linear PRI 07-NOV-2001
LOCUS Homo sapiens parotid secretory protein mRNA, complete cds.
DEFINITION
ACCESSION AF432917
VERSION AF432917.1 GI:16755849
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Venkatesh, S.G., Geetha, C. and Gorr, S.-U.
A member of the PSP/plunc family of BPI proteins is expressed in
the human parotid gland

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
SUBMITTED (12-OCT-2001) Molecular, Cellular and Craniofacial
Biology, University of Louisville, 501 S. Preston Street,
Louisville, KY 40292, USA
Location/Qualifiers
1. .750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/sex="male"
/tissue_type="parotid"
1. .750
/note="salivary protein; putative BPI protein; PSP;
similar to BSP30 and plunc/linx"
/codon_start=1
/product="parotid secretory protein"
/protein_id="AAU28113.1"
/db_xref="GI:16755850"
/translation="MQLMWLVLLCGVLGTSSLLDNLNLSNVVKLEPVLHGL
ETVNTLKGILEKLYDLGWSAQWAKQAQAEKLLNNVTSKLLPTNDFGL
KISNLIIDVKAEPIDDKGNLSFPVTANVTAGPIIGIINLKASLLLTAVRIET
DPQHPQVAVLRCAEDPTGISLSLDDKHSQILNKFNVSINLTKSTVSSLLQKEICP
LIRIFHSLDVNVIOQVDPNPKHTOLOLI"
BASE COUNT 201 a 210 c 163 g 176 t
ORIGIN

Query Match 93.0%; Score 694; DB 9; Length 750;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 744; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATGCTTACAGTTTGAACACTTGTCTCTGTCGGCGTGTCTCACTGGGAGCTCAGAGTCT 61
Db 1 ATGCTTACAGTTTGAACACTTGTCTCTGTCGGCGTGTCTCACTGGGAGCTCAGAGTCT 60

QY 62 CTCTTTGACATCTTGGCAATGACCTAAGCAATGCTGTGGTAAAGCTGGAACTGTCTT 121
Db 61 CTCTTTGACATCTTGGCAATGACCTAAGCAATGCTGTGGTAAAGCTGGAACTGTCTT 120

QY 122 CACGAGGAGCTTGACAGACTTGACAACTACTCTTAAGCGATCTCTTGAGAACTCAAGTCT 181
Db 121 CACGAGGAGCTTGACAGACTTGACAACTACTCTTAAGCGATCTCTTGAGAACTCAAGTCT 180

QY 182 GACCTAGGAGTCTTTCAGAAATCCAGTGTGCAACTTGGCAAGTGGCCAGGAGG 241
Db 181 GACCTAGGAGTCTTTCAGAAATCCAGTGTGCAACTTGGCAAGTGGCCAGGAGG 240

QY 242 GCTGAGAAATGCTGAACAATGCTTCTTAAGCTGCTTCAAGCTGCAAGCTGATGATG 301
Db 241 GCTGAGAAATGCTGAACAATGCTTCTTAAGCTGCTTCAAGCTGCAAGCTGATGATG 300

QY 302 GGGTTGAAATCAGCAACTCCCTCATCTCGATGATCAAGCTGAACCGATCGATGATG 361
Db 301 GGGTTGAAATCAGCAACTCCCTCATCTCGATGATCAAGCTGAACCGATCGATGATG 360

QY 362 AAAGCGCTTAACCTGAGCTTCCCTGTCCAGCGGATGCTGCTGAGCGGCGGCGCATTT 421
Db 361 AAAGCGCTTAACCTGAGCTTCCCTGTCCAGCGGATGCTGCTGAGCGGCGGCGCATTT 420

QY 422 GGCCGATTTATCAACCTGAAAGCTCTTGGACCTCTGACCGGAGTCAACAATTGAACT 481
Db 421 GGCCGATTTATCAACCTGAAAGCTCTTGGACCTCTGACCGGAGTCAACAATTGAACT 480

QY 482 GATCCCCAGACACACAGCCCTGTTGCGTCTGGGAGAAATGCGCCAGTACCACCAACG 541
Db 481 GATCCCCAGACACACAGCCCTGTTGCGTCTGGGAGAAATGCGCCAGTACCACCAACG 540

QY 542 ATCTCAGTTCTTCTGCTGGCAACACAGCGCAATATCAACAAGTTCTGTAATAGCGTG 601
Db 541 ATCTCAGTTCTTCTGCTGGCAACACAGCGCAATATCAACAAGTTCTGTAATAGCGTG 600

QY 602 ATCAACACGCTCAAGACACTGTATCTTCCCTGCTGCAGAGGAGATATCTCCACTGATC 661
Db 601 ATCAACACGCTCAAGACACTGTATCTTCCCTGCTGCAGAGGAGATATGTCCTACTGATC 660

QY 662 CUCATCTTTCATCCACTCCCTGCTGATGTGAATGTCAATTCACGAGGCTGCTGATAATCCCTCAG 721
Db 661 CUCATCTTTCATCCACTCCCTGCTGATGTGAATGTCAATTCACGAGGCTGCTGATAATCCCTCAG 720

QY 722 CACAAACCCAGCTGCMAACCTCA 746
Db 721 CACAAACCCAGCTGCMAACCTCA 745

RESULT 3
HSBA49G10
LOCUS
DEFINITION
Human DNA sequence from clone RP11-49G10 on chromosome 20. Contains a gene similar to bovine salivary protein BSP30, the LOC51297 gene for LUNX protein, 5' end of a gene encoding a protein similar to murine von Ebner minor salivary gland protein, a novel gene, a pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene similar to ribosomal protein L12, a putative novel transcript, ESTs, STSS, GSSs and a CpG island, complete sequence.
ACCESSION AL121901
VERSION AL121901.20
KEYWORDS GI:8249854
SOURCE HTG: BSP30; CpG island; LOC51297; LUNX; RPL12; STAT-inhibitor. human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Tracey, A.
Direct Submission
Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 5, 2000 this sequence version replaced gi:8017404.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
IMPORTANT: This sequence is not the entire insert of clone RP11-49G10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-49G10 is at 1 in this sequence. The true left end of clone RP5-1187J4 is at 161494 in this sequence. The true right end of clone RP4-733023 is at 27823 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-49G10 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers
1. .161593
/organism="Homo sapiens"

FEATURES
source

Category	Feature	Value
misc_feature	misc_feature	48600..48787
	misc_feature	/gene="BA49G10.1"
	misc_feature	/note="match: cDNAs: Em:U79414 Em:U79413 Em:AX061621
	misc_feature	match: ESTs: Em:AI654622"
	misc_feature	/product="BA49G10.1 (similar to bovine salivary protein
	misc_feature	BSP30)"
	misc_feature	/evidence=not_experimental
	misc_feature	35535..48787
	misc_feature	/gene="BA49G10.1"
	misc_feature	join(36522..36678,40308..40452,41455..41562,42783..42935)
repeat_region	repeat_region	45523..45604,46980..47043,47896..47936)
	repeat_region	/gene="BA49G10.1"
	repeat_region	/note="match: proteins: Tr:P79125 Sw:P07743"
	repeat_region	/codon_start=1
	repeat_region	/evidence=not_experimental
	repeat_region	/product="BA49G10.1 (similar to bovine salivary protein
	repeat_region	BSP30)"
	repeat_region	/protein_id="CAC03546.1"
	repeat_region	/db_xref="GI:9801234"
	repeat_region	/translations="MLQLKLVLLCGVLTGTSSLLDNLNDLNVVVKLEPVLHGL
misc_feature	misc_feature	ETVNTLKDKLEKLVGLQSSAQLAKQAERKLNINVISKLLPTNDIFGL
	misc_feature	KISNLILDKAEIPIDGKLNLSFPVANTVAGPIGQIINLKASLDLTTATET
	misc_feature	DPTQHPVAVLGECAVTSISLKLKHSQIINKFVNSVINTLKSTVSSLLQKEICP
	misc_feature	LIRIFHSLDVNVQVVDNPQHKTQLTL"
	misc_feature	38155..38751
	misc_feature	/note="LIP48 repeat: matches 4514..5125 of consensus"
	misc_feature	38750..39785
	misc_feature	/note="LIP48 repeat: matches 5128..6163 of consensus"
	misc_feature	48769..48774
	misc_feature	/gene="BA49G10.1"
misc_feature	misc_feature	48787
	misc_feature	/gene="BA49G10.1"
	misc_feature	52540..53450
	misc_feature	/note="CPG island"
	misc_feature	/evidence=not_experimental
	misc_feature	complement(53831..54131)
	misc_feature	/note="match: GSS: Em:AQ897745"
	misc_feature	53831..54141
	misc_feature	/note="match: STS: Em:G31731"
	misc_feature	53842..53969
misc_feature	misc_feature	/note="match: GSS: Em:AQ429347"
	misc_feature	complement(53844..54141)
	misc_feature	/note="match: GSS: Em:AF101963"
	misc_feature	53858..54022
	misc_feature	/note="match: GSS: Em:AQ392891"
	misc_feature	complement(53908..54141)
	misc_feature	/note="match: GSS: Em:AQ415287"
	misc_feature	53958..54139
	misc_feature	/note="match: GSS: Em:AQ378081"
	misc_feature	53970..54139
misc_feature	misc_feature	/note="match: GSS: Em:AQ527291"
	misc_feature	53972..54142
	misc_feature	/note="match: GSS: Em:AQ429347"
	misc_feature	54000..54141
	misc_feature	/note="match: GSS: Em:AQ392510"
	misc_feature	complement(54027..54139)
	misc_feature	/note="match: GSS: Em:AQ533658"
	misc_feature	54030..54138
	misc_feature	/note="match: GSS: Em:AG000935"
	misc_feature	54041..54149
misc_feature	misc_feature	/note="match: GSS: Em:AQ392891"
	misc_feature	54767..55668
	misc_feature	/gene="BA49G10.2"
	misc_feature	54767..55668
	misc_feature	/gene="BA49G10.2 (similar to STAT-induced STAT
	misc_feature	inhibitor-2)
	misc_feature	match: proteins: Tr:O95102 Tr:O88582 Tr:O14542 Tr:O14509
	misc_feature	Tr:O62225 Tr:O15097 Tr:O35718 Tr:O90X77 Tr:O90X78
	misc_feature	Tr:O70512 Tr:O35960 Tr:O9PW70 Tr:O9Y5R1"
	misc_feature	21.3%; score 159; DB 9; Length 161593;
misc_feature	misc_feature	Best Local Similarity 100.0%; Pred. No. 7.8e-82
	misc_feature	Query Match
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
misc_feature	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
misc_feature	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
misc_feature	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 161593;
	misc_feature	Best Local Similarity
	misc_feature	21.3%; score 159; DB 9; Length 1615

tRNA

```
join(35535. 35585,36507. 36678,40308. 40452,41455. 41562,
42783. 42935,45523. 45604,46980. 47043,47896. 47973,
```

Query Match
Best Local

21.3%; Score 159; DB 9; Length 161593;
100.0%; Pred NO. 7.8e-82.

Matches 159: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 GATGCTTCAGCTTTGAAACTTGTCTCTCTGCGCGCTGCTCACTGGGACCTCAGAGTC 60
 Db 36521 GATGCTTCAGCTTTGAAACTTGTCTCTCTGCGCGCTGCTCACTGGGACCTCAGAGTC 36580
 QY 61 TCTTCTTGACATCTTGGAATGACCTAAGCAATGCTGTGTAAGCTGGAACTGTCTTCT 120
 Db 36581 TCTTCTTGACATCTTGGAATGACCTAAGCAATGCTGTGTAAGCTGGAACTGTCTTCT 36640
 QY 121 TCACGAGGACCTTGACACAGTGTGACAATACTCTTAAAGG 159
 Db 36641 TCACGAGGACCTTGACACAGTGTGACAATACTCTTAAAGG 36679

RESULT 4
 BTU79413
 LOCUS
 DEFINITION Bos taurus common salivary protein BSP30 mRNA, form a, complete cds
 ACCESSION U79413
 VERSION U79413.1 GI:1710366
 KEYWORDS
 SOURCE
 ORGANISM Bos taurus
 Cow
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 999)
 Haigh,B.J., Wilkins,R.J. and Wheeler,T.T.
 The cloning and sequencing of two cDNAs coding for alternate forms of BSP30, a bovine member of the Parotid Secretory Protein family
 Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 999)
 AUTHORS Haigh,B.J., Wilkins,R.J. and Wheeler,T.T.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-1996) Dairy Science Group, AgResearch, Private Bag, Hamilton, New Zealand
 FEATURES
 Source Location/Qualifiers
 1..999
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /tissue_type="parotid gland"
 36..767
 /note="Common salivary protein, form a"
 /codon_start=1
 /product="BSP30"
 /protein_id="AAB38282.1"
 /db_xref="GI:1710367"
 36..767
 CDS
 /note="Common salivary protein, form a"
 /codon_start=1
 /product="BSP30"
 /protein_id="AAB38282.1"
 /db_xref="GI:1710367"
 36..767
 /translation="MVQLKVLVLCGLAGTSESLLDIRGNDVLRLLISGLERCHCTF
 DSTILIFQNLKTELESRCNDVVEETQOTENSLEGLISRIFOVYVRLTGVIRNVQV
 PDITEATSENSANVLPITADVTVSLPLGEIVKLDLNDLQTSVSIETDPTGDSR
 VVVGCPNPSPESILTVLHRRPGLNDVDFGVNLKQLVSSVVQVHELCPRIKELLES
 LDTECKRLGEPQVTQOESEVHR"
 BASE COUNT 267 a 247 c 255 g 230 t
 ORIGIN

Query Match 3.1%; Score 23; DB 4; Length 999;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TTCAGCTTTGGAACCTTGTTCTC 28
 Db 40 TTCAGCTTTGGAACCTTGTTCTC 62

RESULT 5
 BTU79414
 LOCUS
 DEFINITION Bos taurus common salivary protein BSP30 mRNA, form b, complete cds.
 ACCESSION U79414
 VERSION U79414.1 GI:1710366
 KEYWORDS
 SOURCE
 ORGANISM Bos taurus
 Cow
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 1024)
 Haigh,B.J., Wilkins,R.J. and Wheeler,T.T.
 The cloning and sequencing of two cDNAs coding for alternate forms of BSP30, a bovine member of the Parotid Secretory Protein family
 Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 1024)
 AUTHORS Haigh,B.J., Wilkins,R.J. and Wheeler,T.T.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-1996) Dairy Science Group, AgResearch, Private Bag, Hamilton, New Zealand
 FEATURES
 Source Location/Qualifiers
 1..1024
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /tissue_type="parotid gland"
 35..757
 /note="Common salivary protein, form b"
 /codon_start=1
 /product="BSP30"
 /protein_id="AAB38283.1"
 /db_xref="GI:1710369"
 35..757
 CDS
 /note="Common salivary protein, form b"
 /codon_start=1
 /product="BSP30"
 /protein_id="AAB38283.1"
 /db_xref="GI:1710369"
 35..757
 /translation="MVQLKVLVLCGLAGTSESLLDIRGNDVLRLLISGLERGLDTF
 DSTILIFQNLKTELESRCNDVVEETQOTENSLEGLISRIFOVYVRLTGVIRNVQV
 PDITEATSENSANVLPITADVTVSLPLGEIVKLDLNDLQTSVSIETDPTGDSR
 GCTNPSPESILTVLHRRPGLNDVDFGVNLKQLVSSVVQVHELCPRIKELLES
 ECVEKLIGESODTQOEPEGR"
 BASE COUNT 289 a 251 c 254 g 230 t
 ORIGIN

Query Match 3.1%; Score 23; DB 4; Length 1024;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TTCAGCTTTGGAACCTTGTTCTC 28
 Db 39 TTCAGCTTTGGAACCTTGTTCTC 61

RESULT 6
 AK024183/c
 LOCUS Homo sapiens cDNA FLJ14121 fis, clone MAMMAL002009.
 DEFINITION
 ACCESSION AK024183
 VERSION AK024183.1 GI:10436500
 KEYWORDS
 SOURCE Homo sapiens Mammary gland cDNA to mRNA, clone_lib:MAMMAL
 clone:MAMMAL002009.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (sites)
 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
 NEDO human cDNA sequencing project
 Unpublished (2000)
 JOURNAL
 REFERENCE 2 (bases 1 to 1730)
 AUTHORS Isogai,T. and Otsuki,T.
 TITLE Direct Submission
 JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,

Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp,
Tel: 81-438-52-3951, Fax: 81-438-52-3952)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES Location/Qualifiers

1..1730
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAMMA1002009"
/clone_lib="MAMMA1"
/tissue_type="Mammary gland"
/note="cloning vector: pME18SFL3"

BASE COUNT 547 a 358 c 383 g 442 t

ORIGIN

Query Match 2.8%; Score 21; DB 9; Length 1730;

Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 TCAGAAATTCCTGACCAATGT 264

Db 1113 TCAGAAATTCCTGACCAATGT 1093

RESULT 7

HSJ9614 79043 bp DNA linear PRI 17-NOV-2000
LOCUS Homo sapiens chromosome 18 sequence from PAC RPCI-1 34E08, complete
DEFINITION

ACCESSION AJ009614
VERSION G1:11228421
KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 79043)

Authors Radelof, U., Hennig, S., Ramser, J., Francis, F., Steffens, C., Klein, M., Seranski, P., Poustka, A., Reinhardt, R. and Lehrach, H.

Unpublished

2 (bases 1 to 79043)

JOURNAL MPIG.

AUTHORS Direct Submission

TITLE Submitted (17-JUL-1998) MPIG, Abt. Lehrach, Max Planck Institut fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany

On Nov 19, 2000 this sequence version replaced gi:9931108.

contig 01 1..79043

Clone received from the Resource Centre of the Human Genome Project at the Max-Planck-Institute for Molecular Genetics.

COMMENT

FEATURES

source

Location/Qualifiers

1..79043

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/clone_lib="PAC RPCI-1 34E08"

/clone_lib="RPCI1.3-5 Human PAC library, originating institute: Roswell Park Cancer Institute, creator: Pieter de Jong, P. Ioannou"

misc_feature

/note="SP6_end:PAC RPCI-1 34E08"

misc_feature

/note="T7_end:PAC RPCI-1 34E08"

BASE COUNT 22105 a 17090 c 17115 g 22733 t

ORIGIN

Query Match

2.8%; Score 21; DB 9; Length 79043;

Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GGACCTCAGATCTCTCTTCTTG 68

Db 67164 GGACCTCAGATCTCTCTTCTTG 67184

RESULT 8

AC084416_5

WPCOMMENT

Sequence split into 6 fragments LOCUS AC084416 Accession AC084416

Fragment Name Begin End

AC084416_0 1 110000

AC084416_1 100001 210000

AC084416_2 200001 310000

AC084416_3 300001 410000

AC084416_4 400001 510000

AC084416_5 500001 595289

Continuation (6 of 6) of AC084416 from base 500001 (AC084416 Mus musculus clone rp23-

Query Match 2.8%; Score 21; DB 2; Length 95289;

Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 CTTCAGCTTTGGAAACTTGT 25

Db 53091 CTTCAGCTTTGGAAACTTGT 53111

RESULT 9

AC005800

LOCUS

DEFINITION AC005800 100259 bp DNA linear PRI 08-DEC-1998

sequence. Homo sapiens Chromosome 15q26.1 PAC clone pDJ427d15, complete

ACCESSION AC005800

VERSION AC005800.1

KEYWORDS GI:3980456

SOURCE HTG.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 100259)

Authors Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M.,

Buettner, J., Bumelster, R., Card, P., deSailboat, F., Dunn, J.,

English, C., Ethridge, S., Garner, H.R., Gordon, M., Grant, O.,

Hahner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N., Major, T.,

McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J.,

Schultz, R.A., Stimson, S., Wagner, N., Waller, K. and Ward, T.

HTGS Submission

Unpublished

2 (bases 1 to 100259)

Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M.,

Buettner, J., Butler, C., Card, P., deSailboat, F., Dunn, J.,

English, C., Ethridge, S., Garner, H.R., Gordon, M., Grant, O.,

Hahner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N., Major, T.,

McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J.,

Schultz, R.A., Stimson, S., Wagner, N., Waller, K. and Ward, T.

Direct Submission

Submitted (07-OCT-1998) Genome Science & Technology Center,

University of Texas Southwestern Medical Center, 5323 Harry Hines

Bldg, Dallas, TX 75235-8591, USA

3 (bases 1 to 100259)

Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M.,

Buettner, J., Butler, C., Card, P., deSailboat, F., Dunn, J.,

English, C., Ethridge, S., Garner, H.R., Gordon, M., Grant, O.,

Grant, O., Hahner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N.,

Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S.,

Schageman, J., Schultz, R.A., Stimson, S., Waller, K. and Ward, T.

Direct Submission

Submitted (08-DEC-1998) Genome Science & Technology Center,

University of Texas Southwestern Medical Center, 5323 Harry Hines

JOURNAL

```

repeat_region complement(64428, .64702)
               /rpt_family="Alu"
repeat_region complement(65451, .65735)
               /rpt_family="Alu"
repeat_region 66069, .66363
               /rpt_family="Alu"
repeat_region 66511, .66757
               /rpt_family="Alu"
repeat_region 66827, .67040
               /rpt_family="L1"
repeat_region 67136, .67409
               /rpt_family="Alu"
repeat_region 67591, .67869
               /rpt_family="Alu"

```

Query Match 2.8% Score 21; DB 9; Length 100259;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 537 CCAGCATCTCCTTCCTTCG 557
      |||||
Db 3894 CCAGCATCTCCTTCCTTCG 39014

```

RESULT 10
 AC084416.4
 WPCOMMENT

Sequence split into 6 fragments LOCUS AC084416 Accession AC084416

```

Fragment Name      Begin End
AC084416.0         1 110000
AC084416.1        100001 210000
AC084416.2        200001 310000
AC084416.3        300001 410000
AC084416.4        400001 510000
AC084416.5        500001 592289

```

Continuation (5 of 6) of AC084416 from base 400001 (AC084416 Mus musculus clone rp23-430)

Query Match 2.8% Score 21; DB 2; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 5 CTTGAGCTTGGAACTTGTT 25
      |||||
Db 31498 CTTGAGCTTGGAACTTGTT 31518

```

RESULT 11
 AC002312
 LOCUS

DEFINITION Human Chromosome 15 pac pDJ138023, complete sequence.
 AC002312 157248 bp DNA linear PRI 17-DEC-1997
 AC002312.1 GI:2695566

VERSION HTG.
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
 1 (bases 1 to 157248)

AUTHORS Evans, G.A., Athanasiou, M., Basit, M., Bradbury, P., Brignac, S.,
 Bumesiter, R., Davis, C., English, C., Franklin, T.L., Garner, H.R.,
 Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J.,
 Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P.,
 Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
 Ward, T., and Wilson, R.

template

Unpublished

2 (bases 1 to 157248)

REFERENCE Evans, G.A., Bradbury, P., Brignac, S., Bumesiter, R., Burbee, D.,
 Davies, J., Davies, C.J., Davis, C., English, C., Fondon, T.,
 Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O.,
 Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U.,
 Newton, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S.,

Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
 Ward, T., and Wilson, R.

TITLE
 JOURNAL

Direct Submission
 Submitted (09-JUL-1997) Genome Science and Technology Center,
 University of Texas Southwestern Medical Center at Dallas, 5323
 Harry Hines Blvd, Dallas, TX 75235-8591, USA

REFERENCE
 AUTHORS

3 (bases 1 to 157248)
 Evans, G.A., Athanasiou, M., Basit, M., Bradbury, P., Brignac, S.,
 Bumesiter, R., Davis, C., English, C., Franklin, T.L., Garner, H.R.,
 Gee, V., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J.,
 Hinson, S., Narayanaswamy, U., Newton, J., O'Brien, K., Patel, P.,
 Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D.,
 Ward, T., and Wilson, R.

TITLE
 JOURNAL

Direct Submission
 Submitted (17-DEC-1997) Genome Science & Technology Center,
 University of Texas Southwestern Medical Center, 5323 Harry Hines
 Blvd, Dallas, TX 75235-8591, USA

COMMENT On Dec 17, 1997 this sequence version replaced gi:2251215.

FEATURES
 source

Location/Qualifiers
 1..157248
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /clone="pDJ138023"

BASE COUNT 45707 a 33920 c 34072 g 43549 t
 ORIGIN

Query Match 2.8% Score 21; DB 9; Length 157248;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 537 CCAGCATCTCCTTCCTTCG 557
      |||||
Db 18926 CCAGCATCTCCTTCCTTCG 18946

```

RESULT 12
 AC094795/c

DEFINITION Rattus norvegicus clone CH230-4N16, *** SEQUENCING IN PROGRESS ***
 AC094795 158769 bp DNA linear HTG 20-DEC-2001
 AC094795 69 unordered pieces.

VERSION AC094795.2 GI:17941575
 KEYWORDS HTG: HTGS_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae;
 Rattus.
 1 (bases 1 to 158769)

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barbaria, J.,
 Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
 Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., D'achorne, S.R., David, R., Davila, M.L., Davis, C.,
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Louiseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Nextson, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogih, M., Okunoda, G., Oragunye, N., Oviiedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Reby, G., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,
 Sisson, A., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Thomas, N.,
 Tang, H., Tauney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wlecczyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 156769)
 Worley, K.C.
 Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15624631.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: ngsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBJ2
 Center clone name: CH230-4N16
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findphrapblast
 Consensus quality: 120703 bases at least Q40
 Consensus quality: 131098 bases at least Q30
 Consensus quality: 139407 bases at least Q20
 Estimated insert size: 124919; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-tp estimation
 Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 69 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 3820: contig of 3820 bp in length
 * 3821: gap of unknown length
 * 3921: contig of 8884 bp in length
 * 10805: gap of unknown length
 * 10905: contig of 5054 bp in length
 * 15959: gap of unknown length
 * 16059: contig of 4113 bp in length
 * 20172: gap of unknown length
 * 20272: contig of 4212 bp in length
 * 24484: gap of unknown length
 * 24584: contig of 4271 bp in length
 * 28855: gap of unknown length
 * 28955: contig of 2220 bp in length
 * 31175: gap of unknown length
 * 31275: contig of 2945 bp in length
 * 34320: gap of unknown length
 * 34320: contig of 2881 bp in length
 * 37201: gap of unknown length
 * 37301: contig of 2284 bp in length
 * 39585: gap of unknown length
 * 39585: contig of 2911 bp in length
 * 42596: gap of unknown length
 * 42596: contig of 4167 bp in length
 * 46862: gap of unknown length
 * 46862: contig of 2973 bp in length
 * 49935: gap of unknown length
 * 50035: contig of 3320 bp in length
 * 53355: gap of unknown length
 * 53355: contig of 3584 bp in length
 * 57039: gap of unknown length
 * 57039: contig of 3055 bp in length
 * 60394: gap of unknown length
 * 60394: contig of 2867 bp in length
 * 63161: gap of unknown length
 * 63161: contig of 1800 bp in length
 * 65062: gap of unknown length
 * 65062: contig of 2264 bp in length
 * 67425: gap of unknown length
 * 67425: contig of 1456 bp in length
 * 68981: gap of unknown length
 * 68981: contig of 1940 bp in length
 * 71021: gap of unknown length
 * 71021: contig of 3106 bp in length
 * 74327: gap of unknown length
 * 74327: contig of 2079 bp in length
 * 76406: gap of unknown length
 * 76406: contig of 2585 bp in length
 * 79091: gap of unknown length
 * 79091: contig of 2040 bp in length
 * 81231: gap of unknown length
 * 81231: contig of 2321 bp in length
 * 83652: gap of unknown length
 * 83652: contig of 2440 bp in length
 * 86192: gap of unknown length
 * 86192: contig of 2877 bp in length
 * 89169: gap of unknown length
 * 89169: contig of 1759 bp in length
 * 91028: gap of unknown length
 * 91028: contig of 2554 bp in length
 * 93682: gap of unknown length
 * 93682: contig of 2298 bp in length
 * 96080: gap of unknown length
 * 96080: contig of 1267 bp in length
 * 96180: gap of unknown length
 * 96180: contig of 1815 bp in length
 * 97447: gap of unknown length
 * 97447: contig of 1643 bp in length
 * 99362: gap of unknown length
 * 99362: contig of 1580 bp in length
 * 101642: gap of unknown length
 * 101642: contig of 1761 bp in length
 * 103322: gap of unknown length
 * 103322: contig of 1565 bp in length
 * 104987: gap of unknown length
 * 104987: contig of 1088 bp in length
 * 105088: gap of unknown length
 * 105088: contig of 1763 bp in length
 * 106730: gap of unknown length
 * 106730: contig of 1463 bp in length
 * 108209: gap of unknown length
 * 108209: contig of 1601 bp in length
 * 108309: gap of unknown length
 * 108309: contig of 1938 bp in length
 * 110070: gap of unknown length
 * 110070: contig of 2223 bp in length
 * 111879: gap of unknown length
 * 111879: contig of 1088 bp in length
 * 111979: gap of unknown length
 * 111979: contig of 1763 bp in length
 * 113067: gap of unknown length
 * 113067: contig of 1463 bp in length
 * 113167: gap of unknown length
 * 113167: contig of 1601 bp in length
 * 114930: gap of unknown length
 * 114930: contig of 1938 bp in length
 * 115031: gap of unknown length
 * 115031: contig of 2223 bp in length
 * 116494: gap of unknown length
 * 116494: contig of 1088 bp in length
 * 118194: gap of unknown length
 * 118194: contig of 1763 bp in length
 * 118295: gap of unknown length
 * 118295: contig of 1463 bp in length
 * 120333: gap of unknown length
 * 120333: contig of 1601 bp in length
 * 122555: gap of unknown length
 * 122555: contig of 1938 bp in length
 * 122655: gap of unknown length
 * 122655: contig of 2223 bp in length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

```

* 122656 124575: contig of 1920 bp in length
* 124576 124675: gap of unknown length
* 124676 126688: contig of 2013 bp in length
* 126689 126788: gap of unknown length
* 126789 127979: contig of 1191 bp in length
* 127980 128079: gap of unknown length
* 128080 129619: contig of 1540 bp in length
* 129620 129719: gap of unknown length
* 129720 132208: contig of 2489 bp in length
* 132209 132308: gap of unknown length
* 132309 133622: contig of 1314 bp in length
* 133623 133722: gap of unknown length
* 133723 135016: contig of 1294 bp in length
* 135017 135116: gap of unknown length
* 135117 136454: contig of 1338 bp in length
* 136455 136554: gap of unknown length
* 136555 137971: contig of 1417 bp in length
* 137972 138071: gap of unknown length
* 138072 139350: contig of 1279 bp in length
* 139351 139450: gap of unknown length

Query Match      2.8%: Score 21; DB 2: Length 158769;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 671 ATCCACTCCTCGATGTGAAT 691
Db 20726 ATCCACICCTCGATGTGAAT 20706

RESULT 13
AL137059
LOCUS Human DNA sequence from clone Rpl1-125123 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL137059
VERSION AL137059.20 GI:11493221
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 166141)
Lovell, J.
Direct Submission
Submitted (03-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 30, 2000 this sequence version replaced gi:11414453.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rpl1-125123 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.ctori.org/bacpac/home.htm
VECTOR: pBACE3.6

```

```

This sequence is the entire insert of clone Rpl1-125123 The true
left end of clone Rpl1-8910 is at 99174 in this sequence. The true
right end of clone Rpl1-400120 is at 40034 in this sequence.

FEATURES             source
1..166141
    Location/Qualifiers
        Jordanism="Homo sapiens"
        db_xref="taxon:9606"
        chromosome="13"
        clone="Rpl1-125123"
        clone_lib="RPC1-11.1"
7..897
    /note="match: GSS: Em:AQ346776"
repeat_region 577..883
    /note="AluY repeat: matches 1..310 of consensus"
repeat_region 1382..1522
    /note="L2 repeat: matches 1273..1423 of consensus"
repeat_region 1585..1653
    /note="L1MC5 repeat: matches 7729..7797 of consensus"
repeat_region 2289..2330
    /note="5S repeat: matches 1..42 of consensus"
repeat_region 2846..2995
    /note="MIR repeat: matches 102..261 of consensus"
    complement(3929..4189)
    /note="match: GSS: Em:AQ577939"
misc_feature 3945..4203
    /note="match: GSS: Em:AG019010"
misc_feature 3951..4176
    /note="match: GSS: Em:A2517766"
    complement(3951..4071)
misc_feature /note="match: GSS: Em:A2485047"
3954..4148
    /note="match: GSS: Em:A2371182"
    complement(3954..4216)
misc_feature /note="match: GSS: Em:AQ214300"
    complement(3955..4132)
misc_feature /note="match: STS: Em:HS351WB1"
3955..4070
    /note="58 copies 2 mer at 73% conserved"
3957..4181
    /note="match: GSS: Em:AG012444"
3957..4156
    /note="match: STS: Em:G45158"
3959..4102
    /note="match: GSS: Em:A2416994"
    complement(3960..4157)
misc_feature /note="match: STS: Em:G01739"
    complement(3962..4129)
misc_feature /note="match: GSS: Em:AQ985402"
    complement(3962..4102)
misc_feature /note="match: STS: Em:M84924"
    complement(3988..4199)
3995..4144
    /note="match: GSS: Em:A2030802"
    /note="match: GSS: Em:A2009992"
4004..4203
    /note="match: GSS: Em:AQ598220"
    complement(4004..4194)
misc_feature /note="match: GSS: Em:A2353055"
    complement(4020..4153)
misc_feature /note="match: GSS: Em:AQ209289"
    complement(4023..4178)
misc_feature /note="match: GSS: Em:A2306702"
    complement(4040..4210)
misc_feature /note="match: GSS: Em:A2369902"
    complement(4042..4166)
misc_feature /note="match: GSS: Em:A2020968"
    complement(4060..4203)
misc_feature /note="match: GSS: Em:A2116221"
4062..4202
    /note="match: GSS: Em:A2226345"
4063..4194
    /note="match: GSS: Em:A2023541"
4082..4210

```

```

/note="match: GSS: Em:A2023172"
complement(4082..4178)
/note="match: GSS: Em:A2110479"
complement(4088..4203)
/note="match: GSS: Em:B64246"
4130..4203
/note="37 copies 2 mer at 68% conserved"
complement(5247..5696)
/note="match: GSS: Em:AQ127121"
5692..6802
/note="LIMB3 repeat: matches 4280..5471 of consensus"
6803..7111
/note="AluSg repeat: matches 1..313 of consensus"
7112..7812
/note="LIMB3 repeat: matches 5471..6180 of consensus"
8187..8287
/note="MIR repeat: matches 88..197 of consensus"
9286..9727
/note="MSTB repeat: matches 1..426 of consensus"
10159..10271
/note="MIR repeat: matches 32..145 of consensus"
10500..10613
/note="MIR repeat: matches 21..134 of consensus"
10871..11244
/note="MSTB repeat: matches 79..394 of consensus"
11350..11501
/note="MIR repeat: matches 20..172 of consensus"
12922..13067
/note="LIMB3 repeat: matches 6001..6155 of consensus"
13795..13847
/note="LTR16A repeat: matches 359..423 of consensus"
14465..14550
/note="L2 repeat: matches 2611..2702 of consensus"
14590..14965
/note="L2 repeat: matches 2130..2516 of consensus"
15236..15359
/note="FLAM_C repeat: matches 11..132 of consensus"
15736..15983
/note="LIMB1 repeat: matches 5679..5949 of consensus"
16034..16284
/note="LIMB1 repeat: matches 5271..5590 of consensus"
16286..16579
/note="AluSg repeat: matches 1..294 of consensus"
16589..16848
/note="LIMB6 repeat: matches 5867..6172 of consensus"
16849..17009
/note="AluSg/x repeat: matches 137..297 of consensus"
17022..17063
/note="LIM4 repeat: matches 5807..5858 of consensus"
17109..17188
/note="40 copies 2 mer cc 73% conserved"
17140
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 225bp by restriction digest data."
17141..17191
/note="Assembly confirmed by restriction digest.Sequence
from overlapping clone BA400120(AL138599)."
17212..17513
/note="AluSc repeat: matches 1..303 of consensus"
17888..19890
/note="LIMB6 repeat: matches 3790..5813 of consensus"
19891..20236
/note="LIM4 repeat: matches 2932..3283 of consensus"
20408..20571
/note="FRAM repeat: matches 5..173 of consensus"
complement(20443..20909)
/note="match: GSS: Em:AQ562307"
20572..20646
/note="LIM4 repeat: matches 4106..4184 of consensus"
20938..21039
/note="26 copies 2 mer ca 80% conserved"
21041..21195
/note="MIR repeat: matches 107..262 of consensus"

```

```

repeat_region 21450..22001
/note="LIMB3 repeat: matches 5635..6184 of consensus"
22069..22199
/note="MIR repeat: matches 11..144 of consensus"
22820..23059
/note="MIR repeat: matches 5..262 of consensus"
23173..23406
/note="AluSg repeat: matches 20..297 of consensus"
23525..23817
/note="AluJ0 repeat: matches 18..311 of consensus"
23876..24337
/note="LIPAS repeat: matches 5682..6143 of consensus"
24449..24522

Query Match 2.8%; Score 21; DB 9; Length 166141;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 TCCTGGATGTCAAAGCTGAAC 347
|||||
DB 112834 TCCTGGATGTCAAAGCTGAAC 112854

RESULT 14
AC097610
LOCUS
DEFINITION
AC097610 169468 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-163L24, *** SEQUENCING IN PROGRESS
*** 59 unordered pieces.
AC097610
AC097610.3 GI:17973837
HTG: HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 169468)
AUTHORS
Muzny,D.M., Adams,C., Adio-oduola,B., Ali-oshan,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Fallis,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Galis,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

```


* 162136 163228: contig of 1093 bp in length
 * 163229 163328: gap of unknown length
 * 163329 164797: contig of 1469 bp in length
 * 164798 164897: gap of unknown length
 * 164898 165914: contig of 1017 bp in length
 * 165915 166014: gap of unknown length

Query Match 2.8%; Score 21; DB 2: Length 169468;

Best Local Similarity 100.0%; Pred. No. 2.3; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0;

QY 5 CTTGAGCTTTGGAACTTGTT 25
 |||||

Db 75687 CTTGAGCTTTGGAACTTGTT 75707

RESULT 15

AP001854/c

LOCUS

DEFINITION

AP001854

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

109272 125082 contig of 15811 bp in length
 125183 137528 contig of 12346 bp in length
 137629 144522 contig of 6894 bp in length
 144623 150860 contig of 6238 bp in length
 150961 155932 contig of 4972 bp in length
 156033 160622 contig of 4590 bp in length
 160723 164247 contig of 3525 bp in length
 164348 167467 contig of 3120 bp in length
 167568 169824 contig of 2257 bp in length

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 39834: contig of 39834 bp in length
 * 39835 39934: gap of 100 bp
 * 39935 69620: contig of 29686 bp in length
 * 69621 69720: gap of 100 bp
 * 69721 90776: contig of 21056 bp in length
 * 90777 90876: gap of 100 bp
 * 90877 109171: contig of 18295 bp in length
 * 109172 109271: gap of 100 bp
 * 109272 125082: contig of 15811 bp in length
 * 125083 125182: gap of 100 bp
 * 125183 137528: contig of 12346 bp in length
 * 137529 137628: gap of 100 bp
 * 137629 144522: contig of 6894 bp in length
 * 144523 146222: gap of 100 bp
 * 146223 150860: contig of 6238 bp in length
 * 150861 150960: gap of 100 bp
 * 150961 155932: contig of 4972 bp in length
 * 155933 160622: contig of 4590 bp in length
 * 160623 160722: gap of 100 bp
 * 160723 164247: contig of 3525 bp in length
 * 164248 164347: gap of 100 bp
 * 164348 167467: contig of 3120 bp in length
 * 167468 167567: gap of 100 bp
 * 167568 169824: contig of 2257 bp in length.

FEATURES

source

1..169824
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q23"
 /clone="RP11-871M19"

misc_feature

1..39834
 /note="assembly_fragment clone_end:SP6 vector_side:left"

misc_feature

39935..69620
 /note="assembly_fragment clone_end:T7 vector_side:right"

misc_feature

69721..90776
 /note="assembly_fragment"

misc_feature

90877..109171
 /note="assembly_fragment"

misc_feature

109272..125082
 /note="assembly_fragment"

misc_feature

125183..137528
 /note="assembly_fragment"

misc_feature

137629..144522
 /note="assembly_fragment"

misc_feature

144623..150860
 /note="assembly_fragment"

misc_feature

150961..155932
 /note="assembly_fragment"

misc_feature

156033..160622
 /note="assembly_fragment"

misc_feature

160723..164247
 /note="assembly_fragment"

misc_feature

164348..167467
 /note="assembly_fragment"

NOTE: This is a 'working draft' sequence. It currently consists of
 13 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs of N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

1 39834 contig of 39834 bp in length
 39935 69620 contig of 29686 bp in length
 69721 90776 contig of 21056 bp in length
 90877 109171 contig of 18295 bp in length

```

/misc_feature      /note="assembly_fragment"
167568..169824
/note="assembly_fragment"
BASE COUNT      4888 a 35977 c 35308 g 48450 t 1201 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GCACCTCAGACTCTCTCTTC 68
|||||
DB 113934 GCACCTCAGACTCTCTCTTC 113974

RESULT 16
LOCUS      AC067986      171180 bp      DNA      linear      HTG 04-MAR-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-405A15 map 15, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
AC067986
VERSION      AC067986.4 GI:13194985
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 171180)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,I., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castelle,A., Choepey,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lakoque,K., Lamazares,R., Landers,T., Lehoszky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McSheeters,R.,
Meldrum,J., Menkus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,K., Royov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 4, 2001 this sequence version replaced gi:12711782.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9332
Center clone name: 405_A.15
----- Summary Statistics
Sequencing vector: M13; M77815; 4% of reads
Sequencing vector: Plasmid; n/a; 96% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
FEATURES
source
Location/Qualifiers
1..171180
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-405A15"
/clone_lib="RPC1-11 Human Male BAC"
1..744
/note="assembly_fragment"

```

Assembly program: Phrap; version 0.960731
Consensus quality: 161734 bases at least Q40
Consensus quality: 165777 bases at least Q30
Consensus quality: 167633 bases at least Q20
Insert size: 142000; agarose-fp
Quality coverage: 168880; sum-of-contigs
Quality coverage: 5.1 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 744: contig of 744 bp in length
745 844: gap of 100 bp
845 1502: contig of 658 bp in length
1503 1602: gap of 100 bp
1603 3185: contig of 1583 bp in length
3186 3285: gap of 100 bp
3286 5525: contig of 2240 bp in length
5526 5625: gap of 100 bp
5626 7859: contig of 2234 bp in length
7860 7959: gap of 100 bp
7960 9703: contig of 1744 bp in length
9704 9803: gap of 100 bp
9804 12551: contig of 2748 bp in length
12552 12651: gap of 100 bp
12652 14941: contig of 2290 bp in length
14942 15041: gap of 100 bp
15042 17448: contig of 2407 bp in length
17449 17548: gap of 100 bp
17549 20738: contig of 3190 bp in length
20739 20838: gap of 100 bp
20839 23501: contig of 2663 bp in length
23502 23601: gap of 100 bp
23602 26867: contig of 3266 bp in length
26868 26967: gap of 100 bp
26968 30611: contig of 3644 bp in length
30612 30711: gap of 100 bp
30712 35383: contig of 4672 bp in length
35384 35483: gap of 100 bp
35484 40859: contig of 5376 bp in length
40860 40959: gap of 100 bp
40960 85677: contig of 44718 bp in length
85678 85777: gap of 100 bp
85778 93108: contig of 7331 bp in length
93109 93208: gap of 100 bp
93209 99119: contig of 5911 bp in length
99120 99219: gap of 100 bp
99220 110625: contig of 11406 bp in length
110626 110725: gap of 100 bp
110726 120994: contig of 10269 bp in length
120995 121094: gap of 100 bp
121095 131361: contig of 10267 bp in length
131362 131461: gap of 100 bp
131462 146767: contig of 15306 bp in length
146768 146867: gap of 100 bp
146868 162913: contig of 16046 bp in length
162914 163013: gap of 100 bp
163014 171180: contig of 8167 bp in length.


```

misc_feature      clone_end:SP6
vector_side:left
845..1502
/note="assembly_fragment"
misc_feature      1603..3185
/note="assembly_fragment"
3286..3525
/note="assembly_fragment"
5626..7859
/note="assembly_fragment"
7960..9703
/note="assembly_fragment"
9804..12551
/note="assembly_fragment"
12652..14941
/note="assembly_fragment"
15042..17448
/note="assembly_fragment"
17549..20738
/note="assembly_fragment"
20839..23501
/note="assembly_fragment"
23602..26867
/note="assembly_fragment"
26968..30611
/note="assembly_fragment"
30712..35383
/note="assembly_fragment"
35484..40859
/note="assembly_fragment"
40960..485677
/note="assembly_fragment"
85778..93108
/note="assembly_fragment"
93209..99119
/note="assembly_fragment"
99220..110625
/note="assembly_fragment"
110726..120994
/note="assembly_fragment"
121095..131361
/note="assembly_fragment"
131462..146767
/note="assembly_fragment"
146868..162913
/note="assembly_fragment"
163014..171180
/note="assembly_fragment"
clone_end:T7
vector_side:right
BASE COUNT      40509 a 42714 c 44557 g 41074 t 2326 others
ORIGIN

```

```

Query Match      2.8%; Score 21; DR 2; Length 171180;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 CCAGCATCTCACTTCCTTCG 557
|||||
Db 151914 CCAGCATCTCACTTCCTTCG 151894

RESULT 17
AC087675
LOCUS
DEFINITION      Homo sapiens chromosome 18 clone RP11-871M19 map 18, *** SEQUENCING
IN PROGRESS ***; 5 ordered pieces.
ACCESSION      AC087675
VERSION
KEYWORDS
SOURCE
ORGANISM      Homo sapiens

```

```

AC087675
LOCUS
DEFINITION      Homo sapiens chromosome 18 clone RP11-871M19 map 18, *** SEQUENCING
IN PROGRESS ***; 5 ordered pieces.
ACCESSION      AC087675
VERSION
KEYWORDS
SOURCE
ORGANISM      Homo sapiens

```

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo; 1 (bases 1 to 172628)

Birren,B., Linton,L., Nusbaum,C. and Lander,E. 1999. The human genome: initial mapping and sequencing. *Science* 286: 2054-2058.

Homo sapiens chromosome 18, clone RP11-871M19

Unpublished

2 (bases 1 to 172628)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boughey,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choquet,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., DeArillano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glue,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Harford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Reita,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 11, 2002 this sequence version replaced gi:12229450.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12254

Center clone name: 871_M_19

TITLE

JOURNAL

COMMENT

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 109535 109634: gap of 100 bp

* 109635 122255: contig of 12621 bp in length

* 122256 122355: gap of 100 bp

* 123156 125178: contig of 2823 bp in length

* 125179 125278: gap of 100 bp

* 125279 143152: contig of 17874 bp in length

* 143153 143252: gap of 100 bp

* 143253 172628: contig of 29376 bp in length.

Location/Qualifiers

1. 172628

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-871M19"

/clone.lib="RPC1-11 Human Male BAC"

FEATURES

source

BASE COUNT 48931 a 36776 c 36345 g 50098 t 478 others

ORIGIN

Query Match 2.8% Score 21; DB 2; Length 172628;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 48 GGACCTCAGAGTCTCTTTG 68
 |||||
 DB 114256 GGACCTCAGAGTCTCTTTG 114276

RESULT 18

AC106518
 LOCUS
 DEFINITION AC106518 183553 bp DNA linear HTG 12-JAN-2002
 Rattus norvegicus clone CH230-30L6, WORKING DRAFT SEQUENCE, 55
 unordered pieces.

AC106518

AC106518.1 GI:18139042

HTG: HTGS.PHASE1: HTGS_DRAFT.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 183553)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Banks, T., Barbara, J.,

Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J.,

Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Cartron, T.F.,

Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,

Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C., Cox, C.,

Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,

Dean, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,

Duan, R., Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,

Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,

Foster, P., Franco, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,

Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,

Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,

Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,

Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,

Joudav, C., Karlsson, E., Kelly, S., Khan, J., King, L., Korvah, J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,

Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,

Louis, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,

Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,

Mei, G., Melzer, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,

Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,

Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,

Oguy, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,

Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,

Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,

Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,

Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,

Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,

Tang, H., Taubey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,

Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,

Washington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,

Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,

Weinstock, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 183553)

Worley, K.C.

Direct Submission

Submitted (12-JAN-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT


```

/clone_lib="RP31"
1..28242
/clone="assembly_fragment
clone_end:SP6
vector_side:right"
28343..101574
/clone="assembly_fragment
clone_end:T7
vector_side:right"
101675..185983
/clone="assembly_fragment"
54691 a 38748 c 39092 g 53241 t 211 others
ORIGIN

Query Match          2.8%; Score 21; DB 2; Length 185983;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTTGAGCTTTGGAACCTGTT 25
|||||
Db 181345 CTTGAGCTTTGGAACCTGTT 181365

RESULT 20
AC013459/c
AC013459
LOCUS          188741 bp      DNA      linear      HTG 20-OCT-2001
DEFINITION    Homo sapiens chromosome 2 clone RP11-169L20, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
ACCESSION     AC013459
VERSION       AC013459.8 GI:16303541
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 188741)
The sequence of Homo sapiens clone
Unpublished
Waterston,R.H.
2 (bases 1 to 188741)
Direct Submission
Waterston,R.H.
Submitted (12-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Oct 20, 2001 this sequence version replaced gi:16259089.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: H_NH0169L20
----- Summary Statistics -----
Sequencing vector: M13; 30%
Chemistry: Dye-terminator; plasmid; 62%
Chemistry: Dye-terminator Big Dye; 62% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 188437 bases at least Q40
Consensus quality: 188864 bases at least Q30
Consensus quality: 189195 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will

/clone="assembly_fragment
clone_end:SP6
vector_side:right"
28343..101574
/clone="assembly_fragment
clone_end:T7
vector_side:right"
101675..185983
/clone="assembly_fragment"
54691 a 38748 c 39092 g 53241 t 211 others
ORIGIN

Query Match          2.8%; Score 21; DB 2; Length 188741;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 TGAGAAATTGCTGAACAATGT 264
|||||
Db 93476 TGAGAAATTGCTGAACAATGT 93456

RESULT 21
AC021144
AC021144
LOCUS          199254 bp      DNA      linear      HTG 07-JUL-2000
DEFINITION    Homo sapiens chromosome 18 clone RP11-451L19, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
ACCESSION     AC021144
VERSION       AC021144.3 GI:7232137
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 199254)
The sequence of Homo sapiens clone
Unpublished
Waterston,R.H.
2 (bases 1 to 199254)
Direct Submission
Waterston,R.H.
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7024036.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0451L19
----- Summary Statistics -----
Sequencing vector: M13; 84%
Chemistry: Dye-terminator; plasmid; 16%
Chemistry: Dye-terminator Big Dye; 16% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 193747 bases at least Q40
Consensus quality: 195776 bases at least Q30
Consensus quality: 196779 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 197954; sum-of-contigs
Quality coverage: 5.33 in Q20 bases; agarose-fp
Quality coverage: 5.45 in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 3590: contig of 3590 bp in length
* 3591 3690: gap of unknown length
* 3691 980: contig of 6290 bp in length
* 980 1080: gap of unknown length
* 1080 14570: contig of 4490 bp in length
* 14570 19682: contig of 5012 bp in length
* 19682 19783: gap of unknown length
* 19783 29782: contig of 10000 bp in length
* 29782 29882: gap of unknown length
* 29882 37507: contig of 7624 bp in length
* 37507 45734: contig of 8128 bp in length
* 45734 45835: gap of unknown length
* 45835 56272: contig of 10438 bp in length
* 56272 56373: gap of unknown length
* 56373 67026: contig of 10654 bp in length
* 67026 67127: gap of unknown length
* 67127 77931: contig of 10704 bp in length
* 77931 87986: contig of 9955 bp in length
* 87986 102302: contig of 14317 bp in length
* 102302 102403: gap of unknown length
* 102403 143860: contig of 41458 bp in length
* 143860 143961: gap of unknown length
* 143961 199254: contig of 55294 bp in length.

```

FEATURES

```

source
1..199254
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/clone="RP11-451L19"
1..3590
/note="assembly_name:Contig4"
3691..9980
/note="assembly_name:Contig5"
10081..14570
/note="assembly_name:Contig6"
14671..19682
/note="assembly_name:Contig7"
19783..29782
/note="assembly_name:Contig8"
29883..37506
/note="assembly_name:Contig9"
37607..45734
/note="assembly_name:Contig10"
45835..56272
/note="assembly_name:Contig11"
56373..67026
/note="assembly_name:Contig12"
67127..77930
/note="assembly_name:Contig13"
77931..87985
/note="assembly_name:Contig14"
87986..102302
/note="assembly_name:Contig15"
102403..143860
/note="assembly_name:Contig16"
143961..199254
/note="assembly_name:Contig17"
clone_end:Sp6
vector_side:left

```

BASE COUNT 54123 a 44847 c 44725 g 54245 t 1314 others
 ORIGIN

Query Match 2.8% Score 21; DB 2; Length 199254;

Best Local Similarity 100.0%; Pred. No. 2.3; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGACCTCAGAGTCTCTCTTG 68

|||||

Db 186953 GGACCTCAGAGTCTCTCTTG 186973

RESULT 22

AC087399

LOCUS

DEFINITION

AC087399

VERSION

AC087399.4 GI:18034731

KEYWORDS

HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.

SOURCE

ORGANISM

Human

REFERENCE

1 (bases 1 to 202032)

AUTHORS

Birtten,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 202032)

AUTHORS

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,

Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,

Cammarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,

Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T.,

Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,

Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,

McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,

Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,

O'Connor,P., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,

Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,

Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,

Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,

Sougnaz,C., Spencer,R., Stange-Thomann,N., Stojanovic,N.,

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,

Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young,G., Zainoun,J.,

Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (28-DEC-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jan 2, 2002 this sequence version replaced gi:16874879.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12076

Center clone name: 451_L_19

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

```

* 1 68280: contig of 68280 bp in length
* 68281 68380: gap of 100 bp
* 68381 71640: contig of 3260 bp in length
* 71641 71740: gap of 100 bp
* 71741 124778: contig of 53038 bp in length
* 124779 124878: gap of 100 bp
* 124879 133231: contig of 8353 bp in length
* 133232 133331: gap of 100 bp
* 133332 177275: contig of 43944 bp in length
* 177276 177375: gap of 100 bp
* 177376 192942: contig of 15567 bp in length
* 192943 193042: gap of 100 bp
* 193043 202032: contig of 8990 bp in length.
FEATURES
    source
        1..202032
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="18"
            /map="18"
            /clone="RP11-451L19"
            /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 54951 a 47220 c 44171 g 54695 t 995 others
ORIGIN

Query Match 2.8%; Score 21; DB 2; Length 202032;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GGACCTCAGACTCTCTCTTG 68
|||||
DB 43010 GGACCTCAGACTCTCTCTTG 43030

RESULT 23
SSP101
LOCUS SSP101 4692 bp mRNA linear MAM 08-APR-1997
DEFINITION S.scrofa mRNA for p101 protein.
ACCESSION Y10742
VERSION Y10742.1 GI:1935001
KEYWORDS p101 protein.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 4692)
AUTHORS Stephens,L.R., Equino,A., Erdjument-Bromage,H., Lui,M., Cooke,F.,
Coadwell,J., Smrcka,A., Thelen,M., Cadwallader,K., Tempst,P. and
Hawkins,P.T.
TITLE The G beta gamma sensitivity of a p13K is dependent upon a tightly
associated adaptor, p101
JOURNAL Cell 89 (1), 105-114 (1997)
MEDLINE 97248491
REFERENCE 2 (bases 1 to 4692)
AUTHORS Stephens,L.R.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-1997) L.R. Stephens, Babraham Institute,
Babraham, Cambridge, CB2 4AT, UK
FEATURES
    source
        1..4692
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /dev_stage="adult"
            /function="regulatory subunit of G-beta-gamma sensitive
            p13K"
            /codon_start=1
            /product="p101 protein"
            /protein_id="CA471730.1"
            /db_xref="GI:1935002"
            /db_xref="SWISS-PROT:O02696"
            /translation="MQGATTCEDTQHIALERCLHGLSLSRSTWSAGLCNCWSL
            QELVSRDPGHFLLEQLQKTRVEQKGYDLLAPLALLFYVLTCTPHPPDSDL

```

```

LKAARTYHRELTPWVPVYCSICQELLTFIDAEKAPGISYORLVRAEQLGSLSTRHSST
VTVLLANPVEVQAEFLDADKLTSTPGSPHSAYITLLHAFQATFCAHCDLSLHRL
QSKTLAELEAIFTETAQELASGIGDAARQWLRTKLOAAGEKAGFGVLDLTKPG
KLRTPIPARCYTYSWNSDFDILKEQLLEQELLOPELDEDEDEDEDEDEDEDA
DGHCAERDSVLSCTSAASHASTLSASSQSGPTLSQRLTSFVSGLSGVDSGYMED
IEESAYERPRPGCHERRHRRPQXFNRIYKFKSTQWLVLRDRSRSLSEGPSGPP
LRRAGSLCPLDSHTLPSPRAQGRSLPOPKLSPQLPGWLLAPASHRRRRPFLSGDE
DPKASTLUVVFGSDRISGKAVRAYSNLRLENRRPLLTFFKLOFFYVPVAKRSRGTG
TPTSPARSOTPPPLTDAPRHGPAELGAAPWBEESNDISHYLGMLDPAYERNVLCIM
HLPPEVLCQSLKAEPRLEGSQAQLDILAMLLYCYCRFAARPVLLQVYQTELTFTTGE
KTTEIFHLSLELHSAATRAIKASGSGKRLGIDGDEAVPLTLQIIYSKGAISGRSR
WSNMEKLTSVNLISKACQOEELDSSTALTNLNTEVVRQTPEKSKGNQISTSGIK
VDKVOIIGSNPCFVCLDQDKILQSVIRCEVSPCYKPEKSSSLCPFPQRPSPYPPAP
ATPOLCSLLCLPIMTFSGALP"
BASE COUNT 916 a 1560 c 1393 g 823 t
ORIGIN

Query Match 2.7%; Score 20; DB 4; Length 4692;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTCGGA 684
|||||
DB 2448 ATCTTCATCCACTCCCTCGGA 2467

RESULT 24
AR026672
LOCUS AR026672 4692 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5856132.
ACCESSION AR026672
VERSION AR026672.1 GI:5937512
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4692)
AUTHORS Stephens,L., Hawkins,P.,Thomas., and Braselmann,S.
TITLE G-beta-gamma regulated phosphatidylinositol-3' kinase
JOURNAL Patent: US 5856132-A 1 05-JAN-1999;
FEATURES
    Location/Qualifiers
        source
            1..4692
            /organism="unknown"
BASE COUNT 916 a 1559 c 1394 g 823 t
ORIGIN

Query Match 2.7%; Score 20; DB 6; Length 4692;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTCGGA 684
|||||
DB 2448 ATCTTCATCCACTCCCTCGGA 2467

RESULT 25
AR026678
LOCUS AR026678 4692 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5856133.
ACCESSION AR026678
VERSION AR026678.1 GI:5937518
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4692)
AUTHORS Stephens,L. and Hawkins,P.,Thomas.
TITLE G-beta-gamma regulated phosphatidylinositol-3' kinase
JOURNAL Patent: US 5856133-A 1 05-JAN-1999;
FEATURES
    Location/Qualifiers
        source
            1..4692

```

```

BASE COUNT      916 a 1559 c 1394 g      823 t
ORIGIN
/organism="unknown"

Query Match      2.7%; Score 20; DB 6; Length 4692;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTGGG 684
|||||
Db 2448 ATCTTCATCCACTCCCTGGG 2467

RESULT 26
LOCUS      AR029061      4692 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5859201.
ACCESSION  AR029061
VERSION     AR029061.1 GI:5941034
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 4692)
AUTHORS     Stephens,L. and Hawkins,P.Thomas.
TITLE       G-beta-gamma-regulated phosphatidylinositol-3' kinase
JOURNAL     Patent: US 5859201-A 1 12-JAN-1999;
FEATURES    Location/Qualifiers
             source
               1..4692
               /organism="unknown"

BASE COUNT      916 a 1559 c 1394 g      823 t
ORIGIN

Query Match      2.7%; Score 20; DB 6; Length 4692;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTGGG 684
|||||
Db 2448 ATCTTCATCCACTCCCTGGG 2467

RESULT 27
LOCUS      AR033965      4692 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5869271.
ACCESSION  AR033965
VERSION     AR033965.1 GI:5949570
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 4692)
AUTHORS     Stephens,L. and Hawkins,P.Thomas.
TITLE       G-beta-gamma-regulated phosphatidylinositol-3' kinase
JOURNAL     Patent: US 5869271-A 1 09-FEB-1999;
FEATURES    Location/Qualifiers
             source
               1..4692
               /organism="unknown"

BASE COUNT      916 a 1559 c 1394 g      823 t
ORIGIN

Query Match      2.7%; Score 20; DB 6; Length 4692;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTGGG 684
|||||
Db 2448 ATCTTCATCCACTCCCTGGG 2467

BASE COUNT      916 a 1559 c 1394 g      823 t
ORIGIN

Query Match      2.7%; Score 20; DB 6; Length 4692;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTGGG 684
|||||
Db 2448 ATCTTCATCCACTCCCTGGG 2467

```

```

RESULT 28
LOCUS      AL607106/c      36777 bp      DNA      linear      PRI 30-NOV-2001
DEFINITION Human DNA sequence from clone RP11-207N22 on chromosome 1, complete
sequence.
ACCESSION  AL607106
VERSION     AL607106.10 GI:17381409
KEYWORDS   HTG.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 36777)
AUTHORS     Johnson,C.
TITLE       Direct Submission
JOURNAL     Submitted (30-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 5, 2001 this sequence version replaced gi:17066057.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emi, EMBL; Swi,
SWISSPROT; Tr, TREMBL; Wpi, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/
This sequence was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-207N22 is from the library RP11-207N22 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-207N22 it may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-26F12 is at 34778 in this sequence.
The true right end of clone RP4-561P1 is at 2000 in this sequence.

FEATURES    Location/Qualifiers
             source
               1..36777
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="1"
               /clone="RP11-207N22"
               /clone_lib="RP11-11.1"

BASE COUNT      10470 a 8380 c 7835 g 10092 t
ORIGIN

Query Match      2.7%; Score 20; DB 9; Length 36777;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 CTCCTGCTGCAGAGGAGA 647
|||||
Db 32048 CTCCTGCTGCAGAGGAGA 32029

RESULT 29
AC107373/c

```

```

LOCUS      AC107373              97192 bp      DNA      linear      HTG 19-JAN-2002
DEFINITION Homo sapiens chromosome 8 clone CTD-2168K21 map 8, WORKING DRAFT
SEQUENCE   SEQUENCE, 3 ordered pieces.
ACCESSION  AC107373
VERSION    AC107373.1  GI:18250049
KEYWORDS   HTG; HTGS-PHASE2; HTGS-DRAFT; HTGS-FULLTOP.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 97192)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckgalter,B.,
            Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Headford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kanat,A., Karatas,A., Kells,C., Lacroque,K.,
            Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
            Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (19-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -----
            Genomic Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submission@genome.wi.mit.edu
            -----
            Project Information
            Center project name: L21408
            Center clone name: 2168_K_21
            -----
            Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 9837 bases at least Q40
            Consensus quality: 96150 bases at least Q30
            Consensus quality: 96306 bases at least Q20
            Insert size: 97000; agarose-fp
            Quality coverage: 14.4 in Q20 bases; agarose-fp
            Quality coverage: 14.4 in Q20 bases; sum-of-contigs
            -----
            NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *
            1      688: contig of 688 bp in length

```

```

*      689 788: gap of 100 bp
*      789 16413: contig of 15625 bp in length
*      16414 16513: gap of 100 bp
*      16514 97192: contig of 80679 bp in length.
FEATURES             Location/Qualifiers
     source           1..97192
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="CTD-2168K21"
                     /clone_lib="CITD Human BAC"
     misc_feature     1..688
                     /note="assembly_fragment"
     misc_feature     789..16413
                     /note="assembly_fragment"
     misc_feature     16514..97192
                     /note="assembly_fragment"
BASE COUNT  27970 a 20026 c 20485 g 28511 t    200 others
ORIGIN
Query Match      2.7%  Score 20;  DB 2;  Length 97192;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 560 GACAAACACAGCCAAATCAT 579
      |||||
Db 20387 GACAAACACAGCCAAATCAT 20368
      |||||

RESULT 30
AF176680
LOCUS      Homo sapiens chromosome 8 clone CTD-2168K21 map 8p21, WORKING DRAFT
DEFINITION Homo sapiens chromosome 8 clone CTD-2168K21 map 8p21, WORKING DRAFT
SEQUENCE   SEQUENCE, 10 unordered pieces.
ACCESSION  AF176680.4  GI:16356849
VERSION    AF176680
KEYWORDS   HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 102781)
AUTHORS   Wen,G., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
            Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schudy,A.,
            Siddiqui,R., Taudien,S., Schilhabel,M.B., Rosenthal,A. and
            Platzer,M.
            Chromosome 8 genomic sequence
            Unpublished
            2 (bases 1 to 102781)
            Polley,A., Wen,G., Baumgart,C., Dette,M., Jahn,N., Schilhabel,M.,
            Menzel,U. and Rosenthal,A.
            Direct Submission
            Submitted (09-AUG-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
            On Oct 24, 2001 this sequence version replaced gi:14325018.
            -----
            Genomic Center
            Center: Insitute of Molecular Biotechnology
            Center code: IMB
            Web site: http://genome.imb-jena.de/
            Contact: gscj-submit@genome.imb-jena.de
            -----
            Project Information
            Center project name: H230
            Center clone name: CTD-2168K21
            -----
            Summary Statistics
            Sequencing vector: M13; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 92759 bases at least Q40
            Consensus quality: 95158 bases at least Q30
            Consensus quality: 96720 bases at least Q20
            Quality coverage: 6.97 x in Q20 bases; sum-of-contigs

```



```

*****
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality 10.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
*****
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1355: contig of 1355 bp in length
1356 1455: gap of unknown length
1456 47105: contig of 45650 bp in length
47106 47205: gap of unknown length
47206 73104: contig of 25899 bp in length
73105 73204: gap of unknown length
73205 83092: contig of 9888 bp in length
83093 83192: gap of unknown length
83193 86863: contig of 3671 bp in length
86864 86963: gap of unknown length
86964 92440: contig of 5476 bp in length
92441 92539: gap of unknown length
92540 93645: contig of 1106 bp in length
93646 94913: contig of 1168 bp in length
94914 95014: gap of unknown length
95014 96079: contig of 1066 bp in length
96080 96179: gap of unknown length
96180 102781: contig of 6602 bp in length.
*****
FEATURES             Location/Qualifiers
     source            1..102781
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="8"
                        /map="8p21"
                        /clone="CTD-2168K21"
BASE COUNT      29216 a 21550 c 20909 g 30206 t    900 others
ORIGIN

Query Match      2.7%; Score 20; DB 2: Length 102781;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 560 GACAAACACAGCCCAATCAT 579
      |||||
Db 41929 GACAAACACAGCCCAATCAT 41948

RESULT 31
AC095654
LOCUS
DEFINITION      Rattus norvegicus clone CH230-8E12, *** SEQUENCING IN PROGRESS ***,
59 unordered pieces.
*****
ACCESSION      AC095654
VERSION        AC095654.3 GI:17942214
KEYWORDS       HTG: HTGS_PHASE1.
SOURCE         Norway rat
ORGANISM       Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 114599)
REFERENCE
AUTHORS        Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbaria, J.,
Benton, J., Bimberg, K., Blankenburg, K., Bounin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,
Burke, P., Burrell, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Chen, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Dean, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgato, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulvyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vesquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, C., Williamson, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 114599)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15799378.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- project information
Center project name: GCXW
Center clone name: CH230-8E12
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 87795 bases at least Q40
Consensus quality: 93599 bases at least Q30
Consensus quality: 98734 bases at least Q20
Estimated insert size: 8335; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation

```

Quality coverage: lx in Q20 bases: sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 4959 1958: contig of 4958 bp in length
* 5059 5058: gap of unknown length
* 10111 10211: contig of 5153 bp in length
* 10212 10311: gap of unknown length
* 10312 12957: contig of 2646 bp in length
* 12958 13057: gap of unknown length
* 13058 16502: contig of 3445 bp in length
* 16503 16602: gap of unknown length
* 16603 18954: contig of 2352 bp in length
* 18955 19054: gap of unknown length
* 19055 2127: contig of 3073 bp in length
* 2128 2227: gap of unknown length
* 2228 24794: contig of 2567 bp in length
* 24795 24894: gap of unknown length
* 24895 28404: contig of 3510 bp in length
* 28405 28504: gap of unknown length
* 28505 30491: contig of 1987 bp in length
* 30492 30591: gap of unknown length
* 30592 33551: contig of 2960 bp in length
* 33552 33651: gap of unknown length
* 33652 35384: contig of 1733 bp in length
* 35385 35484: gap of unknown length
* 35485 38246: contig of 2762 bp in length
* 38247 38346: gap of unknown length
* 38347 40571: contig of 2224 bp in length
* 40571 40670: gap of unknown length
* 40671 41750: contig of 1080 bp in length
* 41751 41850: gap of unknown length
* 41851 43866: contig of 2016 bp in length
* 43867 43966: gap of unknown length
* 43967 43957: contig of 1991 bp in length
* 43958 46057: gap of unknown length
* 46058 47381: contig of 1324 bp in length
* 47382 47481: gap of unknown length
* 47482 49908: contig of 2427 bp in length
* 49909 50008: gap of unknown length
* 50009 51675: contig of 1667 bp in length
* 51676 51775: gap of unknown length
* 51776 54070: contig of 2295 bp in length
* 54071 54170: gap of unknown length
* 54171 56025: contig of 1855 bp in length
* 56026 56125: gap of unknown length
* 56126 58021: contig of 1896 bp in length
* 58022 58121: gap of unknown length
* 58122 60356: contig of 2235 bp in length
* 60357 60456: gap of unknown length
* 60457 62291: contig of 1835 bp in length
* 62292 62391: gap of unknown length
* 62392 64017: contig of 1626 bp in length
* 64018 64117: gap of unknown length
* 64118 66192: contig of 2075 bp in length
* 66193 66292: gap of unknown length
* 66293 68384: contig of 2092 bp in length
* 68385 68484: gap of unknown length
* 68485 69997: contig of 1513 bp in length
* 69998 70097: gap of unknown length
* 70099 71138: contig of 1041 bp in length
* 71139 71239: gap of unknown length
* 71240 73201: contig of 1963 bp in length
* 73202 73501: gap of unknown length
* 73502 74959: contig of 1658 bp in length
* 73302

* 74960 75059: gap of unknown length
* 75060 76788: contig of 1729 bp in length
* 76789 76888: gap of unknown length
* 76889 78614: contig of 1726 bp in length
* 78615 78714: gap of unknown length
* 78715 80028: contig of 1314 bp in length
* 80029 80128: gap of unknown length
* 80129 81607: contig of 1479 bp in length
* 81608 81707: gap of unknown length
* 81708 82980: contig of 1273 bp in length
* 82981 83080: gap of unknown length
* 83081 84451: contig of 1371 bp in length
* 84452 84551: gap of unknown length
* 84552 85619: contig of 1068 bp in length
* 85620 85719: gap of unknown length
* 85720 86846: contig of 1127 bp in length
* 86847 86946: gap of unknown length
* 86947 88116: contig of 1170 bp in length
* 88117 88216: gap of unknown length
* 88217 89583: contig of 1467 bp in length
* 89584 89783: gap of unknown length
* 89784 90906: contig of 1123 bp in length
* 90907 91006: gap of unknown length
* 91007 92384: contig of 1378 bp in length
* 92385 92484: gap of unknown length
* 92485 93963: contig of 1479 bp in length
* 93964 94063: gap of unknown length
* 94064 95219: contig of 1156 bp in length
* 95220 95319: gap of unknown length
* 95320 96927: contig of 1608 bp in length
* 96928 97027: gap of unknown length
* 97028 98328: contig of 1301 bp in length
* 98329 98428: gap of unknown length
* 98429 99667: contig of 1239 bp in length
* 99668 99767: gap of unknown length
* 99768 101053: contig of 1286 bp in length
* 101054 101154: gap of unknown length
* 101154 102282: contig of 1129 bp in length
* 102283 102382: gap of unknown length
* 102383 103416: contig of 1034 bp in length
* 103417 103516: gap of unknown length
* 103517 104373: contig of 1057 bp in length
* 104374 104574: gap of unknown length
* 104574 105936: contig of 1263 bp in length
* 105937 106037: gap of unknown length
* 106037 107408: contig of 1372 bp in length
* 107409 107508: gap of unknown length
* 107509 108514: contig of 1006 bp in length
* 108515 108614: gap of unknown length
* 108615 110166: contig of 1552 bp in length
* 110167 110266: gap of unknown length

Query Match 2.7% Score 20; DB 2; Length 114599;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 GACATTTTGGGTTGAAAT 312

Db 90283 GACATTTTGGGTTGAAAT 90302

RESULT 32
AC008622/C

LOCUS

DEFINITION

AC008622

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AC008622 Homo sapiens chromosome 19 clone CTB-147N14, complete sequence.
DEFINITION Homo sapiens chromosome 19 clone CTB-147N14, complete sequence.
ACCESSION AC008622

VERSION AC008622.5

KEYWORDS HTG.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 134882)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 134882)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134882)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (31-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 31, 2001 this sequence version replaced gi:7711307.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.8.

FEATURES
source

1. .134882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTB-147N14"
BASE COUNT 30770 a 34938 c 35575 g 33599 t
ORIGIN

Query Match 2.7%; Score 20; DB 9; Length 134882;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 TAACCTGAGCTCCCTGTCA 389
|||||
Db 23000 TAACCTGAGCTCCCTGTCA 22981

RESULT 33
AL161454/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL161454 144536 bp DNA linear PRI 13-JUN-2001
Human DNA sequence from clone RP11-72B4 on chromosome 9, complete sequence.
AL161454
AL161454.10 GI:14456186
HTG.
human.
Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 144536)
Direct Submission
Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jun 14, 2001 this sequence version replaced gi:14269905.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,

SWISSPROT: Tr., TREMBL: Wp.; WORMPEP: Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-72B4 is from the library RP11-11.1 constructed by the group of
Pleter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-72B4 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-72B4 is at 1 in this sequence. The
true left end of clone RP11-18014 is at 142537 in this sequence.
The true right end of clone RP11-346B7 is at 81734 in this
sequence.

FEATURES
source

1. 144536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-72B4"
/clone_lib="RP11-11.1"
repeat_region 1935..2764
/note="L2 repeat: matches 1852..2745 of consensus"
repeat_region 3353..3419
/note="Alu repeat: matches 233..298 of consensus"
repeat_region 4106..4149
/note="MIR repeat: matches 103..146 of consensus"
repeat_region 5694..5993
/note="AluX repeat: matches 1..302 of consensus"
repeat_region 6069..6317
/note="MIR repeat: matches 15..262 of consensus"
repeat_region 7021..7128
/note="MLTLD repeat: matches 1..184 of consensus"
repeat_region 7130..7864
/note="LIMC4 repeat: matches 6822..7608 of consensus"
repeat_region 8159..8304
/note="LIMC4 repeat: matches 7813..7952 of consensus"
repeat_region 8324..8359
/note="18 copies 2 mer ac 97% conserved"
repeat_region 8622..8704
/note="MER47 repeat: matches 2242..2323 of consensus"
repeat_region 8702..8757
/note="MER47 repeat: matches 2265..2320 of consensus"
repeat_region 9384..9430
/note="LIM4 repeat: matches 5276..5321 of consensus"
repeat_region 9431..9784
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 9785..9897
/note="LIM4 repeat: matches 5160..5276 of consensus"
repeat_region 9848..10382
/note="MLTLC repeat: matches 20..466 of consensus"
repeat_region 10411..10623
/note="LIM4 repeat: matches 3504..3744 of consensus"
repeat_region 10651..10777
/note="AluB repeat: matches 84..206 of consensus"
repeat_region 10794..11175
/note="MLTLD repeat: matches 1..417 of consensus"
repeat_region 11735..112861
/note="LIP3 repeat: matches 4256..5387 of consensus"
repeat_region 12854..13614
/note="LIP3 repeat: matches 5359..6142 of consensus"
repeat_region 14647..15016
/note="THE1C repeat: matches 1..371 of consensus"
repeat_region 15137..15290
/note="MER5A repeat: matches 9..178 of consensus"
repeat_region 15915..16173
/note="LIME2 repeat: matches 5913..6162 of consensus"
repeat_region 17298..17542
/note="MIR repeat: matches 15..262 of consensus"

```

misc_feature
/note="single clone region. Assembly confirmed by
restriction digest data."
repeat_region
/note="60 copies 2 mer tt 60% conserved"
repeat_region
/note="Alu repeat: matches 1. .296 of consensus"
repeat_region
/note="L1 repeat: matches 5893. .6175 of consensus"
repeat_region
/note="L1 repeat: matches 3985. .4104 of consensus"
repeat_region
/note="AluX repeat: matches 64. .302 of consensus"
repeat_region
/note="L1 repeat: matches 3596. .3984 of consensus"
repeat_region
/note="AluSc repeat: matches 1. .301 of consensus"
repeat_region
/note="L1 repeat: matches 2138. .3596 of consensus"
repeat_region
/note="L1 repeat: matches 1910. .2104 of consensus"
repeat_region
/note="L1 repeat: matches 1489. .1687 of consensus"
repeat_region
/note="AluSg1 repeat: matches 3. .296 of consensus"
repeat_region
/note="L1 repeat: matches 844. .1489 of consensus"
repeat_region
/note="AluX repeat: matches 1. .239 of consensus"
repeat_region
/note="TIGER1 repeat: matches 1. .1792 of consensus"
repeat_region
/note="AluY repeat: matches 1. .293 of consensus"
repeat_region
/note="L1 repeat: matches 2510. .2574 of consensus"
repeat_region
/note="MER1C repeat: matches 1. .1071 of consensus"
repeat_region
/note="L1 repeat: matches 5173. .6178 of consensus"
repeat_region
/note="MER5A repeat: matches 46. .159 of consensus"
repeat_region
/note="L1 repeat: matches 21. .189 of consensus"
repeat_region
/note="L1 repeat: matches 2552. .2809 of consensus"
repeat_region
/note="L1 repeat: matches 2773. .2896 of consensus"
repeat_region
/note="L1 repeat: matches 179. .1438 of consensus"
repeat_region
/note="L1 repeat: matches 1761. .2326 of consensus"
repeat_region
/note="L1 repeat: matches 2405. .3590 of consensus"
repeat_region
/note="L1 repeat: matches 5374. .6224 of consensus"
repeat_region
/note="MIR repeat: matches 177. .256 of consensus"
repeat_region
/note="AluX repeat: matches 1. .304 of consensus"
repeat_region
/note="MIR repeat: matches 22. .177 of consensus"
repeat_region
/note="L1 repeat: matches 4487. .4585 of consensus"
repeat_region
/note="AluSg repeat: matches 11. .294 of consensus"
repeat_region
/note="HAL repeat: matches 414. .1081 of consensus"
repeat_region
/note="HAL repeat: matches 883. .939 of consensus"
repeat_region
/note="L1 repeat: matches 5193. .6224 of consensus"
repeat_region
/note="AluSc repeat: matches 1. .287 of consensus"

```

```

repeat_region
/note="TIGER2 repeat: matches 1451. .2718 of consensus"
repeat_region
/note="AluX repeat: matches 1. .291 of consensus"
repeat_region
/note="AluX repeat: matches 1. .298 of consensus"
repeat_region
/note="L1 repeat: matches 6570. .6692 of consensus"
repeat_region
/note="L1 repeat: matches 4937. .5258 of consensus"
repeat_region
/note="AluSc repeat: matches 1. .290 of consensus"
repeat_region
/note="L1 repeat: matches 5258. .6004 of consensus"
repeat_region
/note="MER3 repeat: matches 17. .49 of consensus"
repeat_region
/note="L1 repeat: matches 7031. .7419 of consensus"
repeat_region
/note="L1 repeat: matches 7261. .7565 of consensus"
repeat_region
/note="L1 repeat: matches 7699. .7827 of consensus"

```

Query Match 2.7% Score 20; DB 9; Length 144536;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 TTCTAAGCTGCTTCCAACT 286
 |||||

Db 100870 TTCTAAGCTGCTTCCAACT 100851

RESULT 34

AC027709/c

LOCUS

DEFINITION

SEQUENCE

AC027709

AC027709.2

HTGS_PHASE1

HTGS_DRAFT

human

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 158276)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 158276)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Gallagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lenocky,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mianga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL

Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Apr 28, 2000 this sequence version replaced gi:7382520.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/HM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7018

Center clone name: 492_P-7

----- Summary Statistics

Sequencing vector: M13, M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 139569 bases at least Q40

Consensus quality: 148371 bases at least Q30

Consensus quality: 151798 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 154276; sum-of-contigs

Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 121: contig of 121 bp in length

* 122 221: gap of 100 bp

* 222 1570: contig of 1349 bp in length

* 1571 1670: gap of 100 bp

* 1671 3170: contig of 1500 bp in length

* 3171 3270: gap of 100 bp

* 3271 4301: contig of 1031 bp in length

* 4302 4401: gap of 100 bp

* 4402 6057: contig of 1656 bp in length

* 6058 6157: gap of 100 bp

* 6158 7234: contig of 1077 bp in length

* 7235 7334: gap of 100 bp

* 7335 8454: contig of 1120 bp in length

* 8455 8554: gap of 100 bp

* 8555 9563: contig of 1009 bp in length

* 9564 9663: gap of 100 bp

* 9664 10803: contig of 1140 bp in length

* 10804 10903: gap of 100 bp

* 10904 12572: contig of 1669 bp in length

* 12573 12672: gap of 100 bp

* 12673 14413: contig of 1741 bp in length

* 14414 14513: gap of 100 bp

* 14514 16204: contig of 1691 bp in length

* 16205 16304: gap of 100 bp

* 16305 18380: contig of 2076 bp in length

* 18381 18480: gap of 100 bp

* 18481 20703: contig of 2223 bp in length

* 20704 20803: gap of 100 bp

* 20804 23199: contig of 2396 bp in length

* 23200 23299: gap of 100 bp

* 23300 26457: contig of 3158 bp in length

* 26458 26557: gap of 100 bp

* 26558 29851: contig of 3294 bp in length

* 29852 29951: gap of 100 bp

* 29952 33128: contig of 3377 bp in length

* 33129 33228: gap of 100 bp

* 33229 34969: contig of 1741 bp in length

* 34970 35069: gap of 100 bp

* 35070 39244: contig of 4175 bp in length

* 39245 39344: gap of 100 bp

* 39345 41990: contig of 2646 bp in length

* 41991 42090: gap of 100 bp

* 42091 45777: contig of 3687 bp in length

* 45778 45877: gap of 100 bp

* 45878 49591: contig of 3714 bp in length

* 49592 49691: gap of 100 bp

* 49692 53508: contig of 3817 bp in length

* 53509 53608: gap of 100 bp

* 53609 57512: contig of 3904 bp in length

* 57513 57612: gap of 100 bp

* 57613 63118: contig of 5506 bp in length

* 63119 63218: gap of 100 bp

* 63219 67091: contig of 3873 bp in length

* 67092 67191: gap of 100 bp

* 67192 71729: contig of 4538 bp in length

* 71730 71829: gap of 100 bp

* 71830 77477: contig of 5648 bp in length

* 77478 77577: gap of 100 bp

* 77578 83123: contig of 5546 bp in length

* 83124 83223: gap of 100 bp

* 83224 89369: contig of 6146 bp in length

* 89370 89469: gap of 100 bp

* 89470 94947: contig of 5478 bp in length

* 94948 95047: gap of 100 bp

* 95048 99326: contig of 4179 bp in length

* 99327 99327: gap of 100 bp

* 99327 105096: contig of 5770 bp in length

* 105097 105196: gap of 100 bp

* 105197 110543: contig of 5347 bp in length

* 110544 110643: gap of 100 bp

* 110644 117212: contig of 6569 bp in length

* 117213 117312: gap of 100 bp

* 117313 123367: contig of 6055 bp in length

* 123368 123467: gap of 100 bp

* 123468 130252: contig of 6785 bp in length

* 130253 130352: gap of 100 bp

* 130353 138602: contig of 8250 bp in length

* 138603 138702: gap of 100 bp

* 138703 146693: contig of 7991 bp in length

* 146694 146793: gap of 100 bp

* 146794 158276: contig of 11483 bp in length.

FEATURES

source

1..158276
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19"
 /clone_lib="RP11-492P7"
 /clone="RP11-492P7" Human Male BAC"

misc_feature

1..121
 /note="assembly_fragment"
 clone_end:17
 vector_side:right
 222..1570

misc_feature

/note="assembly_fragment"
 1671..3170

misc_feature

/note="assembly_fragment"
 3271..4301

misc_feature

/note="assembly_fragment"
 4402..6057

misc_feature

/note="assembly_fragment"
 6158..7234

misc_feature

/note="assembly_fragment"
 7335..8454

misc_feature

/note="assembly_fragment"
 8555..9563

misc_feature

/note="assembly_fragment"
 9664..10803

misc_feature

/note="assembly_fragment"
 10904..12572

misc_feature

/note="assembly_fragment"
 12673..14413

misc_feature

/note="assembly_fragment"
 14514..16204

```

/note="assembly_fragment"
16305..18380
/note="assembly_fragment"
18481..20703
/note="assembly_fragment"
20804..23199
/note="assembly_fragment"
23300..28457
/note="assembly_fragment"
26558..29851
/note="assembly_fragment"
24952..33128
/note="assembly_fragment"
33229..34969
/note="assembly_fragment"

Query Match      2.7%  Score 20;  DB 2;  Length 158276;
Best local Similarity 100.0%;  Pred. No. 8.9;
Matches 20;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 370 TAACCTGAGCTTCCTGTCA 389
|||||
DB 39702 TAACCTGAGCTTCCTGTCA 39683

RESULT 35
AC094124/C
LOCUS
DEFINITION
AC094124 175335 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-3C23, *** SEQUENCING IN PROGRESS ***
69 unordered pieces.
AC094124
AC094124.2 GI:17940834
HTG: HTGS.PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 175335)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Aisbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Ruhay,C.,
Burche,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Deigado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisseg,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Mettaker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

```

Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

AUTHORS
TITLE
JOURNAL

Worley, K.C.
Direct Submission
Submitted (25-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 176601)
Worley, K.C.
Direct Submission
Submitted (01-JUN-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

4 (bases 1 to 176601)
Worley, K.C.
Direct Submission
Submitted (05-JUN-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 176601)
Worley, K.C.
Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jun 1, 2001 this sequence version replaced gi:14190566.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc.help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING HEAD COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 176601
Phrap values in estimate: 175784
Average error rate (BCM-Phrap estimate): 2.0537e-06
Fraction of Phrap values less than 40 : 0.0018261
Number of consensus changing edits: 13
Number of N's in consensus : 0

----- Consensus changing edits ----- Edited+Context
Position Original+Context
17690 atactatgca(n)ccataaaaaa
17999 ttgtgcacat(n)taacttaaaa
30777 gaaggcagg(n)aaataaaaaa
39183 ttgtgtgtt(n)agatggagtc
61309 tactcttat(n)tgctctttn
61317 atctgctctt(n)tnntcttctt
61319 ntgctcttnt(n)ntctctctt
61320 tgctctttn(n)tnntcttctt
61323 tctctnctn(n)tcctctttaa
61462 tacagaaaa(n)caaaatttag
97888 tttttttt(n)gagacggagt
107632 aaaaaaaa(g)gaaacaaaa
175482 atataatttg(n)attatata

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
5001									
4501									
4001									
3501									
3001									
2501									
2001									
1501									
1001									
501			*	*	*	*	*	*	*
01			*	*	*	*	*	*	*
Phrap Value Range									

FEATURES	Version:	1.01	qxfo.
Source	Location/Qualifiers	1..176601	
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/Chromosome="8"		
	/Clones="RP11-42902"		
repeat_region	1632..1681		
repeat_region	/rpt_family="(TG)n"		
repeat_region	1778..1835		
repeat_region	/rpt_family="MIR"		
repeat_region	3291..3324		
repeat_region	/rpt_family="(GGCA)n"		
repeat_region	5588..5818		
repeat_region	/rpt_family="MIR"		
repeat_region	6129..6325		
repeat_region	/rpt_family="MIR"		
repeat_region	6823..7237		
repeat_region	/rpt_family="L1MEC"		
repeat_region	complement(7423..7631)		
repeat_region	/rpt_family="MER58A"		
repeat_region	7789..8075		
repeat_region	/rpt_family="L1MEC"		
repeat_region	8076..8205		
repeat_region	/rpt_family="FLAM_A"		
repeat_region	8206..8768		
repeat_region	/rpt_family="L1MEC"		
repeat_region	8799..8947		
repeat_region	/rpt_family="AluJo/FRAM"		
repeat_region	8948..9001		
repeat_region	/rpt_family="(GGGA)n"		
repeat_region	9009..9275		
repeat_region	/rpt_family="L1MEC"		
repeat_region	complement(9653..9686)		
repeat_region	/rpt_family="MIR"		
repeat_region	complement(9746..10048)		
repeat_region	/rpt_family="Alusg"		


```

repeat_region complement(10061..10213)

Query Match
Best Local Similarity 2.7%; Score 20; DB 9; Length 176601;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 560 GACAAACACAGCCAAATCAT 579
|||||
Db 16229 GACAAACACAGCCAAATCAT 16248

RESULT 37
AC083892/c 190390 bp DNA linear ROD 29-JAN-2002
LOCUS Mus musculus chromosome 1 clone rp23-116m12, complete sequence.
DEFINITION
AC083892
VERSION AC083892.17 GI:17861021
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 190390)
AUTHORS Shaull,S., Rahhal,R., Yao,Z. and Roe,B.A.
JOURNAL Mus musculus Chromosome 1 BAC Clone rp23-116m12
TITLE Unpublished
AUTHORS 2 (bases 1 to 190390)
JOURNAL Shaull,S., Rahhal,R., Yao,Z. and Roe,B.A.
AUTHORS Direct Submission
TITLE Submitted (05-OCT-2000) Department Of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 3 (bases 1 to 190390)
JOURNAL Shaull,S., Rahhal,R., Yao,Z. and Roe,B.A.
AUTHORS Direct Submission
TITLE Submitted (15-DEC-2001) Department Of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 4 (bases 1 to 190390)
JOURNAL Shaull,S., Rahhal,R., Yao,Z. and Roe,B.A.
AUTHORS Direct Submission
TITLE Submitted (29-JAN-2002) Department Of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Dec 15, 2001 this sequence version replaced gi:16152276.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
Source
Location/Qualifiers
1..190390
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-116m12"
/cdate="2001-01-29"
BASE COUNT 52972 a 43000 c 42141 g 52277 t
ORIGIN

Query Match
Best Local Similarity 2.7%; Score 20; DB 10; Length 190390;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 561 ACAACACAGCCAAATCATC 580
|||||
Db 55503 ACAACACAGCCAAATCATC 55484

RESULT 38
AL645842
LOCUS Mus musculus chromosome 11 clone RP23-20A9, *** SEQUENCING IN
DEFINITION

```

```

PROGRESS ***, in unordered pieces.
AL645842
VERSION AL645842.10 GI:17402343
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (sites)
TITLE Sims,S
JOURNAL Direct Submission
COMMENT Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 6, 2001 this sequence version replaced gi:17384553.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM20A9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 205222 bases at least Q40
Consensus quality: 206077 bases at least Q30
Consensus quality: 206730 bases at least Q20
Insert size: 207564; sum-of-contigs
Insert size: 206517; 2.7% error; agarose-fp
Quality coverage: 10.30x in Q20 bases; sum-of-contigs Quality
Coverage: 10.49x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
-----
FEATURES
Source
Location/Qualifiers
1..208964
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-20A9"
/clone_lib="RPCI-23"
/misc_feature 1..37841
/note="assembly_fragment:02912
fragment_chain:1
clone_end:Sp6
vector_side:left"
37942..52689
/note="assembly_fragment:04132
fragment_chain:1"
52790..64009
/note="assembly_fragment:00560
fragment_chain:2"
64110..89188
/note="assembly_fragment:03474
fragment_chain:2"
89289..91329
/note="assembly_fragment:01118
fragment_chain:3"
91430..102272
/note="assembly_fragment:01996
fragment_chain:3"
102373..123288
/note="assembly_fragment:00103
fragment_chain:3"
123389..137701
/note="assembly_fragment:02287
fragment_chain:3"
137802..147257
/note="assembly_fragment:01389

```

```

misc_feature      fragment_chain:3"
147358. .149656
/Note="assembly_fragment:02510
fragment_chain:3"
misc_feature      149757. .161256
/Note="assembly_fragment:02185
fragment_chain:3"
misc_feature      161357. .183703
/Note="assembly_fragment:03987
fragment_chain:3"
misc_feature      183804. .197669
/Note="assembly_fragment:02217
fragment_chain:3"
misc_feature      197770. .200436
/Note="assembly_fragment:04243
fragment_chain:3"
misc_feature      200537. .208964
/Note="assembly_fragment:04156
fragment_chain:3
clone_end:17
vector_side:right"

```

```

BASE COUNT      57230 a 49399 c 47648 g 53281 t 1406 others
ORIGIN

```

```

Query Match      2.7% Score 20; DB 2; Length 208964;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 617 AGCACTGTATCCTCCCTGCT 636
|||||
DB 99247 AGCACTGTATCCTCCCTGCT 99266

```

```

RESULT 39
AC012250/c
LOCUS
DEFINITION      Homo sapiens clone RP11-16P4, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
AC012250
VERSION         AC012250.3 GI:7381808
KEYWORDS        HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE          human.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galsagan,J., Gardyna,S., Grant,G., Hoges,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:6453955.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

FEATURES

```

Source
1. .210995
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-16P4"
/clone_lib="RP11-11 Human Male BAC"
misc_feature      1..1978
/Note="assembly_fragment"
misc_feature      2079..4482
/Note="assembly_fragment"
misc_feature      4583..10713
/Note="assembly_fragment"
misc_feature      10814..18605
/Note="assembly_fragment"
misc_feature      18706..35662
/Note="assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature      35763..57031
/Note="assembly_fragment"
clone_end:17
vector_side:right"
misc_feature      57132..87266
/Note="assembly_fragment"
misc_feature      87367..119483
/Note="assembly_fragment"
misc_feature      119584..155940
/Note="assembly_fragment"

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WILBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3611
Center clone name: 16_P_4
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 204827 bases at least Q40
Consensus quality: 208297 bases at least Q30
Consensus quality: 209308 bases at least Q20
Insert size: 205000; agarose-fp
Quality coverage: 6.0 in Q20 bases; agarose-fp
Quality coverage: 5.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1978: contig of 1978 bp in length
1979 2078: gap of 100 bp
2079 4482: contig of 2404 bp in length
4483 4582: gap of 100 bp
4583 10713: contig of 6131 bp in length
10714 10813: gap of 100 bp
10814 18605: contig of 7792 bp in length
18606 18705: gap of 100 bp
18706 35662: contig of 16957 bp in length
35663 35762: gap of 100 bp
35763 57031: contig of 21269 bp in length
57032 57131: gap of 100 bp
57132 87266: contig of 30135 bp in length
87267 87366: gap of 100 bp
87367 119483: contig of 32117 bp in length
119484 119583: gap of 100 bp
119584 155940: contig of 36357 bp in length
155941 156040: gap of 100 bp
156041 210995: contig of 54955 bp in length.

```

FEATURES

```

Location/Qualifiers
1. .210995
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-16P4"
/clone_lib="RP11-11 Human Male BAC"
misc_feature      1..1978
/Note="assembly_fragment"
misc_feature      2079..4482
/Note="assembly_fragment"
misc_feature      4583..10713
/Note="assembly_fragment"
misc_feature      10814..18605
/Note="assembly_fragment"
misc_feature      18706..35662
/Note="assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature      35763..57031
/Note="assembly_fragment"
clone_end:17
vector_side:right"
misc_feature      57132..87266
/Note="assembly_fragment"
misc_feature      87367..119483
/Note="assembly_fragment"
misc_feature      119584..155940
/Note="assembly_fragment"

```

```

misc_feature      /note="assembly_fragment"
156041..210995
/note="assembly_fragment"
BASE COUNT 57929 a 47946 c 47269 g 56951 t 900 others
ORIGIN
Query Match      2.7%; Score 20; DB 2; Length 210995;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 CTCCTGCTGCAGAGGAGA 647
|||||
Db 179068 CTCCTGCTGCAGAGGAGA 179049

RESULT 40
AL662927/c AL662927 216038 bp DNA linear HTG 30-JAN-2002
LOCUS Homo sapiens chromosome 1 clone RP11-16P4, ** SEQUENCING IN
DEFINITION PROGRESS ***, 2 unordered pieces.
ACCESSION AL662927
VERSION AL662927.8 GI:18250578
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Direct Submission
Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jan 19, 2002 this sequence version replaced g1:18181773.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA16P4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 215543 bases at least Q40
Consensus quality: 215698 bases at least Q30
Consensus quality: 215815 bases at least Q20
Insert size: 215938; sum-of-contigs
Quality coverage: 13.45x in Q20 bases; sum-of-contigs Quality
coverage: 14.75x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 209571: contig of 209571 bp in length
* 209572 209671: gap of 100 bp
* 209672 216039: contig of 6367 bp in length.
FEATURES
Location/Qualifiers
1..216038
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-16P4"
/clone_lib="RP11-1"
misc_feature 1..209571
/note="assembly_fragment.06618

```

```

clone_end:SP6
vector_side:left
clone_end:T7
vector_side:left"
209672..216038
/note="assembly_fragment:01526"
101 others
BASE COUNT 61925 a 50140 c 47530 g 56342 t
ORIGIN
Query Match      2.7%; Score 20; DB 2; Length 216038;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 CTCCTGCTGCAGAGGAGA 647
|||||
Db 126458 CTCCTGCTGCAGAGGAGA 126439

RESULT 41
AC073693
LOCUS AC073693 230372 bp DNA linear HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-152L22, WORKING DRAFT SEQUENCE, 19
unordered pieces.
ACCESSION AC073693
VERSION AC073693.1 GI:8810310
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 230372)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 230372)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
----- Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: http://www.jgi.doe.gov
----- Project Information
Center Project Name: 1795526
Center clone name: RPCL-23_152L22
----- Summary Statistics
Consensus quality: 217356 bases at least Q40
Consensus quality: 223517 bases at least Q30
Consensus quality: 224719 bases at least Q20
Estimated insert size: 258000; agarose-fp estimation
Estimated insert size: 228572; sum-of-contigs estimation
Quality coverage: 7.65 in Q20 bases; agarose-fp estimation
Quality coverage: 8.64 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1087: contig of 1087 bp in length
* 1088 1187: gap of unknown length
* 1188 2715: contig of 1528 bp in length
* 2716 2815: gap of unknown length
* 2816 4138: contig of 1343 bp in length
* 4139 4258: gap of unknown length
* 4259 6474: contig of 2215 bp in length
* 6474 6573: gap of unknown length

```

```

* 6574 8982: contig of 2409 bp in length
* 8983 9082: gap of unknown length
* 9083 11186: contig of 2104 bp in length
* 11187 11288: gap of unknown length
* 11287 17000: contig of 5714 bp in length
* 17001 25179: gap of unknown length
* 25180 34875: gap of unknown length
* 34876 34976: gap of unknown length
* 34976 43664: contig of 8689 bp in length
* 43665 53788: gap of unknown length
* 53789 65520: contig of 11632 bp in length
* 65521 80284: gap of unknown length
* 80285 80384: gap of unknown length
* 80385 98449: contig of 18065 bp in length
* 98450 116262: gap of unknown length
* 116263 140706: gap of unknown length
* 140707 167609: contig of 24344 bp in length
* 167610 167709: gap of unknown length
* 167710 193399: gap of unknown length
* 193400 230372: contig of 31689 bp in length
* 230373 230372: contig of 30874 bp in length.

```

FEATURES

```

source
1..230372
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RPC1 mouse BAC library 23"
/clone_lib="RPC1 mouse BAC library 23"
BASE COUNT 61266 a 53740 c 52522 g 61042 t 1802 others
ORIGIN

```

Query Match 2.78; Score 20; DB 2: Length 230372;

Best Local Similarity 100.0%; Pred. No. 9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 12 TTGGAAACTTCTCTCTG 31
|||||
Db 601:7 TTGGAAACTTCTCTCTG 60166

```

RESULT 42

AC079485/c

LOCUS

DEFINITION

AC079485 240842 bp DNA linear HTG 02-SEP-2000

Mus musculus clone RP23-167E16, WORKING DRAFT SEQUENCE, 57

unordered pieces.

AC079485

AC079485.1 GI:9964850

HTG: HTGS_PHASE1; HTGS_DRAFT.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 240842)

DOE Joint Genome Institute.

Sequencing of Mouse

Unpublished

2 (bases 1 to 240842)

DOE Joint Genome Institute.

Direct Submission

Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 1801112

Center clone name: RPC1-23_167E16

Summary Statistics

Consensus quality: 184430 bases at least Q40

Consensus quality: 208605 bases at least Q30

Consensus quality: 214713 bases at least Q20

Estimated insert size: 203000; agarose-fp estimation

Quality coverage: 9.19 in Q20 bases; agarose-fp estimation

Quality coverage: 7.93 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 57 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1053: contig of 1053 bp in length

* 1054 1153: gap of unknown length

* 1154 2454: contig of 1301 bp in length

* 2455 2554: gap of unknown length

* 2555 4548: contig of 1994 bp in length

* 4549 4648: gap of unknown length

* 4649 5927: contig of 1279 bp in length

* 5928 6027: gap of unknown length

* 6028 7240: contig of 1213 bp in length

* 7241 7340: gap of unknown length

* 7341 8404: contig of 1064 bp in length

* 8405 8504: gap of unknown length

* 8505 9540: contig of 1036 bp in length

* 9541 9640: gap of unknown length

* 9641 11135: contig of 1495 bp in length

* 11136 11235: gap of unknown length

* 11236 12533: contig of 1298 bp in length

* 12534 12633: gap of unknown length

* 12634 14107: contig of 1474 bp in length

* 14108 14207: gap of unknown length

* 14208 15762: contig of 1555 bp in length

* 15763 15862: gap of unknown length

* 15863 17777: contig of 1915 bp in length

* 17778 17877: gap of unknown length

* 17878 18048: contig of 1171 bp in length

* 18049 19149: gap of unknown length

* 19149 20170: contig of 1022 bp in length

* 20171 20270: gap of unknown length

* 20271 21737: contig of 1467 bp in length

* 21738 21837: gap of unknown length

* 21838 23471: contig of 1634 bp in length

* 23472 23571: gap of unknown length

* 23572 24592: contig of 1021 bp in length

* 24593 24692: gap of unknown length

* 24693 26260: contig of 1568 bp in length

* 26261 26360: gap of unknown length

* 26361 27984: contig of 1624 bp in length

* 27985 28084: gap of unknown length

* 28085 29351: contig of 1267 bp in length

* 29352 29451: gap of unknown length

* 29452 32023: contig of 2572 bp in length

* 32024 32123: gap of unknown length

* 32124 33235: contig of 1112 bp in length

* 33236 33335: gap of unknown length

* 33336 35354: contig of 2019 bp in length

* 35355 35454: gap of unknown length

* 35455 37328: contig of 1874 bp in length

* 37329 37428: gap of unknown length

* 37429 39075: contig of 1647 bp in length

* 39076 39175: gap of unknown length

* 39176 40313: contig of 1138 bp in length

* 40314 40413: gap of unknown length

40414 42073: contig of 1660 bp in length
 42074 42173: gap of unknown length
 42174 43008: contig of 1735 bp in length
 43009 43008: gap of unknown length
 43009 43661: contig of 1653 bp in length
 43662 45761: gap of unknown length
 45762 47410: contig of 1649 bp in length
 47411 47510: gap of unknown length
 47511 49396: contig of 1886 bp in length
 49397 49496: gap of unknown length
 49497 50964: contig of 1468 bp in length
 50965 51064: gap of unknown length
 51065 52096: contig of 1032 bp in length
 52097 52196: gap of unknown length
 52197 54065: contig of 1869 bp in length
 54066 54165: gap of unknown length
 54166 56639: contig of 2474 bp in length
 56640 56739: gap of unknown length
 56740 58433: contig of 1594 bp in length
 58434 58533: gap of unknown length
 58534 60561: contig of 2028 bp in length
 60562 60661: gap of unknown length
 60662 62199: contig of 1538 bp in length
 62200 63426: contig of 1127 bp in length
 63427 63526: gap of unknown length
 63527 63884: contig of 2358 bp in length
 63885 65984: gap of unknown length
 65985 69376: contig of 3392 bp in length
 69377 74280: contig of 4804 bp in length
 74281 74380: gap of unknown length
 74381 76418: contig of 1938 bp in length
 76419 76855: contig of 3267 bp in length
 76856 79786: gap of unknown length
 79787 83462: contig of 3676 bp in length
 83463 83561: gap of unknown length
 83562 85993: contig of 2332 bp in length
 85994 88017: contig of 2024 bp in length
 88018 89465: contig of 1348 bp in length
 89466 89565: gap of unknown length
 89566 92468: contig of 2903 bp in length
 92469 92568: gap of unknown length
 92569 94373: contig of 1805 bp in length
 94374 94733: gap of unknown length
 94734 97290: contig of 2817 bp in length
 97291 97390: gap of unknown length
 97391 107307: contig of 9917 bp in length
 107308 107407: gap of unknown length
 120563 120563: contig of 13156 bp in length
 120564 120563: gap of unknown length
 120564 132583: contig of 11920 bp in length
 132584 132683: gap of unknown length
 132684 149580: contig of 16897 bp in length
 149581 149680: gap of unknown length
 149681 169738: contig of 20058 bp in length
 169739 169838: gap of unknown length
 169839 240842: contig of 71004 bp in length.

FEATURES

Source

1..240842
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RP23-167E16"
 /clone_lib="RPC1 mouse BAC library 23"
 59363 a 59436 c 58638 g 57580 t 5625 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 2.7%; Score 20; DB 2; Length 240842;

100.0%; Pred. No. 9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 370 TRACCTGAGCTTCCTGTCA 389
 |||||
 Db 187938 TRACCTGAGCTTCCTGTCA 187919
 |||||
 RESULT 43
 S81230 788 bp DNA linear VRL 09-MAY-1996
 LOCUS UL52' helicase-primase complex homolog [phocid herpesvirus type-2
 DEFINITION PHV-2, host-Phoca vitulina, isolate 7848, Genomic, 788 nt].
 ACCESSION S81230
 VERSION S81230.1 GI:1311488
 KEYWORDS
 SOURCE phocid herpesvirus 2 host-Phoca vitulina.
 ORGANISM
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
 REFERENCE 1 (bases 1 to 788)
 AUTHORS Harder,T.C., Harder,M., Vos,H., Kulonen,K., Kennedy-Stoskopf,S.,
 Liess,B., Appel,M.J. and Osterhaus,A.D.
 TITLE Characterization of phocid herpesvirus-1 and -2 as putative alpha-
 and gammaherpesviruses of North American and European pinnipeds
 J. Gen. Virol. 77 (Pt 1), 27-35 (1996)
 MEDLINE 96145132
 REMARK GenBank staff at the National Library of Medicine created this
 entry [NCBI gibbsq 176141] from the original journal article.
 This sequence comes from Fig. 4a.

FEATURES

Source

1..788
 /organism="phocid herpesvirus 2"
 /isolate="7848"
 /db_xref="taxon:47419"
 2..787
 /partial
 /gene="UL52' helicase-primase complex homolog"
 2..787
 /partial
 /gene="UL52' helicase-primase complex homolog"
 /note="This sequence comes from Fig. 4a"
 /codon_start=1
 /protein_id="AAB35973.1"
 /db_xref="GI:1311489"

/translation="NSEPNLMQLQELGLDFAATAPDGRNCTKGDIFPTVHNIPV
 FRCQFLKNRYFIVNADITLTQVAKSTVLLPOTPNWATTLDMQITERIFYKTFEFLSN
 NKDQLOISRIHYEVNVRVPVNIYLDPLGVKGLTHOYDVCLALREDVTOILQ
 LCDVDPOHOVYEFKSSCPLELDKMFQNCSEKILGLRVNLPRTAIVGSEPLI
 TLVKILNRMVAKNPLFIHLCTILDSECLLILEPTIRKCVRI"
 BASE COUNT 221 a 179 c 168 g 220 t

BASE COUNT

ORIGIN

Query Match 2.5%; Score 19; DB 14; Length 788;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 481 TGATCCCCACACACCAG 499
 |||||
 Db 493 TGATCCCCACACACCAG 511
 |||||

RESULT 44

NGPRIOP/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 2582)

NGPRIOP/C 2582 bp DNA linear BCT 03-APR-1997

N.gonorhoeae proline iminopeptidase and OpaA genes.

225461

225461.1 GI:396839

OpaA gene; proline iminopeptidase.

Neisseria gonorrhoeae.

Neisseria gonorrhoeae.

Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;

Neisseria.

1 (bases 1 to 2582)

AUTHORS Albertson,N.H. and Koomey,M.
 TITLE Molecular cloning and characterization of a proline iminopeptidase
 gene from *Neisseria gonorrhoeae*
 JOURNAL Mol. Microbiol. 9 (6), 1203-1211 (1993)
 MEDLINE 95020651
 REFERENCE 2 (bases 1 to 82)
 AUTHORS Bhat,K.S., Gibbs,C.P., Barrera,O., Morrison,S.G., Jahng,F.,
 Stern,A., Kupsch,E.M., Meyer,T.F. and Swanson,J.
 TITLE The opacity proteins of *Neisseria gonorrhoeae* strain MS11 are
 encoded by a family of 11 complete genes
 JOURNAL Mol. Microbiol. 5 (8), 1889-1901 (1991)
 MEDLINE 92114767
 REMARK Erratum:[published erratum appears in Mol Microbiol 1992
 Apr;6(8):1073-6]]
 REFERENCE 3 (bases 1 to 82)
 AUTHORS Bhat,K.S., Gibbs,C.P., Barrera,O., Morrison,S.G., Jahng,F.,
 Stern,A., Kupsch,E.M., Meyer,T.F. and Swanson,J.
 TITLE The opacity proteins of *Neisseria gonorrhoeae* strain MS11 are
 encoded by a family of 11 complete genes
 JOURNAL Mol. Microbiol. 6 (8), 1073-1076 (1992)
 MEDLINE 92261323
 REFERENCE 4 (bases 1 to 2582)
 AUTHORS Albertson,N.H.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-1993) NAN H ALBERTSON, GENERAL AND MARINE
 MICROBIOLOGY, UNIVERSITY OF, GÖTEBORG, CARL SKOTTSSERGS GATA 22,
 GÖTEBORG, 413 19, SWEDEN
 FEATURES
 source
 1..2582
 /organism="Neisseria gonorrhoeae"
 /strain="MS01-IX, derived from MS11"
 /db_xref="taxon:485"
 /clone="pNA4"
 /complement(11..82)
 /codon_start=1
 /transl_table=11
 /product="OpaA"
 /protein_id="CAA80949.1"
 /db_xref="GI:396841"
 /db_xref="SPTREMBL:O51058"
 /translation="MNPARKKPSLLFSSLLFRSAGGK"
 /note="upstream activation site typical for -24 -12
 promoters"
 410..426
 /note="-24 -12 promoter"
 438..500
 493..497
 501..1433
 /gene="pip"
 501..1433
 /gene="pip"
 /codon_start=1
 /transl_table=11
 /product="proline iminopeptidase Pip"
 /protein_id="CAA80948.1"
 /db_xref="GI:396840"
 /db_xref="SWISS-PROT:P42786"
 /translation="MVEIKOPPHSGYQLQVSEHQIYWEESGNPDGVPVFLHGGPGAG
 ASPECGRFPNPDVERIVILIDRCGRSHPYACADNTTDLVADIEKVRMLGIGKWL
 VEGSGWSTLSLAYAGTHPERVKGLVRLGFLCRPSETANLNEAGGVSRIYPRQWQF
 VAPIENRNRLIEYHGLLFQDEVCLSRAKAWDWESILIRFEPGVDEDAYSL
 AIALENLRYFVNGWLQGDKALNNIGKIRHPTIVYGVGYDLCTPMQSAWELSKAFP
 EAEALRYVQAGHCAFPPLDALVQAVEDILPRL"
 907..916
 /gene="pip"
 /note="gonococcal DNA uptake sequence"
 1448..1477
 stem_loop 496 a 568 c 816 g 702 t
 BASE COUNT
 ORIGIN
 2.5%; Score 19; DB 1; Length 2582;
 Query Match

Best Local Similarity 100.0%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 663 GCATCTTCATCCACTCCCT 681
 Db 1141 GCATCTTCATCCACTCCCT 1123
 RESULT 45
 BD008630/c 2957 bp DNA linear PAT 31-JAN-2002
 LOCUS Compositions and methods for the prevention and diagnosis of human
 DEFINITION granulocytic ehrlichiosis.
 ACCESSION BD008630
 VERSION BD008630.1 GI:18637003
 KEYWORDS JP 2001502528-A/6.
 SOURCE unidentified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 2957)
 AUTHORS Fikrig,E., Barthold,S.W., Ijdo,J. and Sun,W.
 TITLE Compositions and methods for the prevention and diagnosis of human
 JOURNAL granulocytic ehrlichiosis
 COMMENT Patent: JP 2001502528-A 6 27-FEB-2001;
 YALE UNIVERSITY
 OS Unidentified
 PN JP 2001502528-A/6
 PD 27-FEB-2001
 PF 30-SEP-1997 JP 1998516827
 PR
 PI EROL FIKRIG, STEPHEN W BARTHOLD, JACOB IJDO, WEI SUN PC
 C12N15/31, C12N15/62, C12N15/70, C12N1/21, C07K14/29, C12Q1/68, PC
 A61K39/02,
 PC A61K39/40, G01N33/577
 CC Strandedness: Single;
 CC Topology: Linear;
 FH Key
 FT CDS Location/Qualifiers
 1..2957
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 874 a 506 c 829 g 747 t 1 others
 ORIGIN
 Query Match 2.5%; Score 19; DB 6; Length 2957;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 310 AATCAGCAACTCCCTCATC 328
 Db 991 AATCAGCAACTCCCTCATC 973
 RESULT 46
 BPLPSBLOC 15389 bp DNA linear BCT 21-AUG-1998
 LOCUS
 DEFINITION *Bordetella pertussis* lipopolysaccharide biosynthesis locus baf
 gene, waaA, waaB & waaB,C,D,E,F,G,H,I,J & L genes.
 ACCESSION X90711
 VERSION X90711.1 GI:992967
 KEYWORDS baf gene; lipopolysaccharide biosynthesis; terminal inverted
 repeat; waaA gene; waaC gene; wlaA gene; wlaB gene; wlaC gene; wlaD
 gene; wlaE gene; wlaF gene; wlaG gene; wlaH gene; wlaI gene; wlaJ
 gene; wlaK gene; wlaL gene.
 SOURCE *Bordetella pertussis*.
 ORGANISM Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Bordetella.
 REFERENCE 1 (bases 1 to 15389)
 AUTHORS Allen,A. and Maskell,D.
 TITLE The identification, cloning and mutagenesis of a genetic locus

required for lipopolysaccharide biosynthesis in Bordetella pertussis
Molecular microbiology. 19 (1), 37-52 (1996)
96419162
2 (bases 1 to 15389)
Maskell,D.J.
Direct Submission
Submitted (10-AUG-1995) D. J. Maskell, Imperial College of Science,
Technology, and Medicine, Dept of Biochemistry, Exhibition Road,
London SW7 2AZ, UK
3 (bases 1 to 15389)
Reeves,P.K., Hobbs,M., Valvano,M.A., Skurnik,M., Whitfield,C.,
Coplin,D., Klena,J., Maskell,D., Raetz,C.R.H. and
Rick,P.D.

Bacterial polysaccharide synthesis and gene nomenclature
Trends Microbiol. 49, 495-503 (1996)
Polysaccharide genes have been named in accordance with the
Bacterial Polysaccharide Gene Nomenclature scheme.

FEATURES

Source

1. .15389
/organism="Bordetella pertussis"
/strain="Bp536"
/db_xref="taxon:520"
/note="LPS biosynthesis locus"
1. .92
/gene="baf"
<1. .92
/gene="baf"
/codon_start=3
/transl_table=11
/protein_id="CAA62242.1"
/db_xref="GI:992968"
/db_xref="SPTREMBL:Q45373"
/translation="FGATPQPTYLDSPVLDGLAALAAQCAPTA"
complement(289. .1575)
/gene="waaA"
complement(289. .1575)
/gene="waaA"

gene

CDS

gene

CDS

/note="gene referred to as kdtA in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62243.1"
/db_xref="GI:992969"
/db_xref="SPTREMBL:Q45374"
/translation="MGRGVYTLAKGLAPLILMMKRRARRACQWELPAPAFCHAG
AKAPAPAPVWVIAVSLGTRRAADPIVQALLERGLPVLLTHTTATGAFGRKLFQAA
IGRQLOQAWLPYDFPGATRRFLARHAPRGGLLMREVPNLLAARAGVPMALVSA
RFSASSLRQAGWLQALREALGLDYLVAQTDDEGALRQAGANAYVTGSLKFUVAL
PEAQLRVGHAWAGATGRPVIALASTREGDAMELEAIGALQAHKAATPRPLILIPRH
PQRTDEAAALQAAAGLAYARRSAGSGEPGPHIDVLLGLDGLCEMPFYAAADVAIVGGS
FARLGGNLTCAAGTPTVIVGPHTFNFKDAARDALAAAGALRAPDARTALDWAQLLL
AEPARQAMSEARAWTAHAGATRRITLDALEDWLS"
complement(1577. .2584)
/gene="waaC"
complement(1577. .2584)
/gene="waaC"
/note="gene referred to as rfaC in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62244.1"
/db_xref="GI:992970"
/db_xref="SPTREMBL:Q45375"
/translation="MPNRILIVRTSSGLDVLHMLPAISDIARYVPTQIDWAAEAF
DTPKHDPVAVNEIKVAHRRWRKAWSEPVQERRALAERLRSYSYDVLDMDGLKSA
WLVRQTVGRHGLDMSRAREPLVSLRYNVRHVEFPQAVVRORLKAALTFCYNVAGS
PDEGLQAFGRQAQADAPAPGRRLHLAADRGYAVIMPSASRDDKLWTEDDWAVFR
LDQDQAGARLLAGNEQEAERFVLMGEGEVLDFMDLTSVAQVLGARDWGLDS
GLTHLSAALGRTTIGIYIRASTPTVPTPLVGPSTASLGDGRGASPREAVLASVEQALAA
A"

gene

2714. .3766

CDS

/gene="wlbA"
2714. .3766
/gene="wlbA"
/note="gene referred to as bplA in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62245.1"
/db_xref="GI:992971"
/db_xref="SPTREMBL:Q45376"
/translation="MSSLPTIDRKIRFGLVGGGRISKNHIGTAHQHGRALAEVIEDT
NPEALQAAEAATGARPESSLDLAOGNADALVATPSGLHPWQAEVQAGRHVYSE
KPMATRWEDCKRMVKACDEAGVLFVYQKRRNATLQLVKKAIEQGRFGRIVYVTVN
FWTRPQYVYDAARWKGWDEAGFMQASHYVDLLDLVGPVSVYATATLARIE
AEDTGAALRWHRGAMGSINVMTLYPONLEGSITILGEKGTVRVGVAVNRIDENKF
AEPHDDDKIREANVETTSYVGCHPLYDINCLRGDCEPETDGRGLQSLALLTA
IYRSARGVRIPPLD"
3770. .4345
/gene="wlbB"
3770. .4345
/gene="wlbB"
/note="gene referred to as bplB in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62246.1"
/db_xref="GI:992972"
/db_xref="SPTREMBL:Q45377"
/translation="MTTHPTAIVDEGARIGANSRIWHVHICGGAIECAGCSLQGV
FVGNVRIGDRVKIONNVSYDNVFLDEDDVFCGSPMVFTNVNPRAAIERKNEYRDTL
VROGATIGANCTIVCGATVGRYAFVGAGAVNKDVPDFALVGVGPAPRIGWMSRHGEQ
LDLPLAGNGQARCPHTGDLYLENGVCRLEQ"
4349. .5449
/gene="wlbC"
4349. .5449
/gene="wlbC"
/note="gene referred to as bplC in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62247.1"
/db_xref="GI:992973"
/db_xref="SPTREMBL:Q45378"
/translation="MQFDLKTQYQALRDTINPR.QAVLDHGGOFIMGPEVKELEAALC
AVTGAKCHITVASGTEALLISLMALGVKAGDEVITTSFTFVATAEVIALGAKPVEVD
VEPTCNKILSEIEAKITPRTKAILPVSLYCGQGDMDDEVAVAAHGLIPVIEDAQS
GATTGRKSCNLSITIGTSFPSPKLCYGGGALFTNDDELACAMREIRVHGQSGRY
YHARGVGRMDTIQCAVVLGKLEPDMETIAQRIKIGARYQQLLADLPGGACTVTRP
DRSVAAQFTVMVNPNEAVIAQLKEAGIPTAVHYPRPIHAQPAYEYAEAGAGATPVSD
DLAARVSLPMHPDLEATQDKIVAALRQALN"
5457. .6545
/gene="wlbD"
5457. .6545
/gene="wlbD"
/note="gene referred to as bplD in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62248.1"
/db_xref="GI:992974"
/db_xref="SPTREMBL:Q45379"
/translation="MPKKILTVLGGARPOFTIKASVWSAAIAQHPELTVVVHTQGHFDA
NMSDFFDELGMQTPAHQLDTHGGHGMGTGRMLVALEQVMAQAEKPDVVLVYGDGTNST
LAGALAAVKLHIPVAHVAEAGLSRNLRMPEEVNRLTDRISRMLFTPTDSATRLAAE
GOSCEQVQGVDMYDVALHGGARVTAEGALAAHGLKPGGYVLATIHRAENTDQAQ
LTTIVRALQALAEORVWPPLPRTWGLARLGLDELASTVTLLEPVGVLDMVQLEK
YAAIATDSQVQKEAEFPHRIPCVTLRDETENTELVDAGNNRLAPPVSSAVQAQVOD
ALRCQPDVQYVGGQARRIPIVDLAAAH"
6616. .7827
/gene="wlbE"
6616. .7827
/gene="wlbE"
/note="gene referred to as bplE in Mol. Microbiol.

```

19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62249.1"
/db_xref="GI:992975"
/db_xref="SPTREMBL:Q45380"
/translation="MEFRPYFGREWHIGHGVKAASTISHIRARAPQAGRLTREN
VDCLEVLVATLPYOCNGARRLLNLFOSARLYGLRRDLGGWRPDIIVASTHPYDVL
PAARLAROTGASLVEVHDLPLTPRLGGFKAMHPMIASQYAEYAYRHADLTVM
LICALPYMERGLDPRYAHVNGVPVTEYSPDEFNDYLYRAQIRQLREQCDFVL
AYAGTHCANALMDLQAMARUDQPIGLLLGDDGPKRKLKAGLGLRHIAFADP
VIRPQVAYMADIDAIYIGLRSPLEFQFVSPNKLFDYMLSACPVQVQIESGNDIVAD
AKGLGSPAEEDFAALAAALHGLRTLPAERQAMGRGRDYVYLARHDYVLAQQFLDVA
QSVTPPRAASR"
gene 7890..9077
CDS 7890..9077
/genes="wlbF"
/notes="gene referred to as bplF in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62250.1"
/db_xref="GI:992975"
/db_xref="SPTREMBL:Q45381"
/translation="WSDFIFFALPDIGEAETQAVTESMRSGWLITGPNAREFERFAA
YIGADVEAVANSATAGLHLALEAIGVGPGEVITHTTASAEVARYLGAEPVLVD
IDPATLCISPAATIERAITPTRAIVPVHYGLSCDMSILEIARKHGLKVEDAAHAL
PASWQGRKSGLESOLTVSYFATKTATGEGMVVTRDPALAKPCRVMLRHGIDROA
FQFTSKKPANVYEIVAPGFKYNMTDTAAAGRVOLQVQOMROHRAQIAAAYDQAF
DPLTLPPCGRTGPGVERVAHRDDEHSHWLYAIRHPQAPLKCDDEIVRMTEGICG
SVHYVPLHLPWRURYGLTPDMYPHSQAAPEGMASLPYISRMITDADVQVRIASVRQL
LRP"
gene 9074..9667
CDS 9074..9667
/genes="wlbG"
/notes="gene referred to as bplG in Mol. Microbiol.
19:37-52(1996)"
/codon_start=1
/transl_table=11
/protein_id="CAA62250.1"
/db_xref="GI:992975"
/db_xref="SPTREMBL:Q45381"
/translation="WSDFIFFALPDIGEAETQAVTESMRSGWLITGPNAREFERFAA
YIGADVEAVANSATAGLHLALEAIGVGPGEVITHTTASAEVARYLGAEPVLVD
IDPATLCISPAATIERAITPTRAIVPVHYGLSCDMSILEIARKHGLKVEDAAHAL
PASWQGRKSGLESOLTVSYFATKTATGEGMVVTRDPALAKPCRVMLRHGIDROA
FQFTSKKPANVYEIVAPGFKYNMTDTAAAGRVOLQVQOMROHRAQIAAAYDQAF
DPLTLPPCGRTGPGVERVAHRDDEHSHWLYAIRHPQAPLKCDDEIVRMTEGICG
SVHYVPLHLPWRURYGLTPDMYPHSQAAPEGMASLPYISRMITDADVQVRIASVRQL
LRP"

```

```

Query Match 2.5%; Score 19; DB 1; Length 15389;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 AACACGACATTTTGGGT 305
|||||
Db 9604 AACACGACATTTTGGGT 9622

RESULT 47
AL670463 30425 bp DNA linear PRI 07-FEB-2002
LOCUS Human DNA sequence from clone Rpl3-145E22 on chromosome X, complete
DEFINITION sequence.
ACCESSION AL670463.3 GI:18642402
VERSION AL670463.3
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 30425)
Chapman, J.
Direct Submission
Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Feb 8, 2002 this sequence version replaced gi:18477112.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep>. This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>. Rpl3-145E22 is from the library RPl3-13.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

FEATURES

Source Location/Qualifiers

1..30425

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone_lib="RPl3-145E22"

/clone_lib="RPl3-13.1"

BASE COUNT 10249 a 5744 c 5296 g 9136 t

ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 30425;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 560 GACAAACACAGCCCAATCA 578

|||||

Db 3140 GACAAACACAGCCCAATCA 3158

RESULT 48

HSB33B7 36339 bp DNA linear PRI 12-DEC-1999

LOCUS Human DNA sequence from clone SC22CB-33B7 on chromosome 22 Contains

DEFINITION GSSs, complete sequence.

ACCESSION 282176

VERSION 282176.1 GI:3550355

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 36339)

McLaren, S.

Direct Submission

Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Sep 8, 1998 this sequence version replaced gi:3452478.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

This sequence was generated from part of bacterial clone contigs of

human chromosome 22, constructed by the Sanger Centre Chromosome 22

Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/c_elegans/wormpep SC22CB-33B7 is from the human chromosome 22-specific cosmid library (SC22CB) constructed at the Sanger Centre by Mark Ross and Cordelia Langford.

VECTOR: lawr1st16

This sequence is the entire insert of clone SC22CB-33B7. The true left end of clone RPL-222E13 is at 25472 in this sequence. The true right end of clone RP5-919B11 is at 27970 in this sequence.

FEATURES

Source	Location/Qualifiers
1. .36339	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="22"
	/clone_lib="SC22CB"
	/clone="SC22CB-33B7"
repeat_region	1. .79
	/note="Alu repeat: matches 234. .312 of consensus"
misc_feature	258. .882
	/note="CpG island"
	/evidence=not_experimental
variation	322. .323
	/note="clone RPL-222E13; gg in this entry; deletion"
	/replace="ggcgcgcgg"
variation	626. .628
	/note="clone RPL-222E13; tag in this entry; substitution"
	/replace="tgg"
repeat_region	659. .806
	/note="74 copies 2 mer gg 57 conserved"
variation	802. .804
	/note="clone RPL-222E13; ggg in this entry; substitution"
	/replace="gcg"
variation	934. .936
	/note="clone RPL-222E13; gcc in this entry; substitution"
	/replace="gtc"
variation	967. .969
	/note="clone RPL-222E13; gat in this entry; substitution"
	/replace="ggt"
repeat_region	985. .1056
	/note="Alu repeat: matches 68. .109 of consensus"
repeat_region	1219. .1535
	/note="AluSc repeat: matches 1. .309 of consensus"
variation	1365. .1367
	/note="clone RPL-222E13; ggc in this entry; substitution"
	/replace="tgc"
variation	1507. .1513
	/note="clone RPL-222E13; aaaaaca in this entry; insertion"
	/replace="aaa"
variation	1594. .1596
	/note="clone RPL-222E13; gtt in this entry; substitution"
	/replace="tgt"
repeat_region	1772. .2078
	/note="AluX repeat: matches 2. .309 of consensus"
variation	1941. .1942
	/note="clone RPL-222E13; tt in this entry; deletion"
	/replace="tat"
variation	2055. .2057
	/note="clone RPL-222E13; tga in this entry; substitution"
	/replace="taa"
repeat_region	2375. .2494
	/note="MIR repeat: matches 33. .167 of consensus"
repeat_region	2617. .2637
	/note="AluX repeat: matches 1. .312 of consensus"
repeat_region	2638. .2946
	/note="MIR repeat: matches 245. .262 of consensus"
	/replace="AluX repeat: matches 1. .307 of consensus"
repeat_region	2947. .3127
	/note="MIR repeat: matches 34. .245 of consensus"
variation	2980. .2982
	/note="clone RPL-222E13; cgt in this entry; substitution"
	/replace="cat"
variation	3136. .3138
	/note="clone RPL-222E13; cgc in this entry; substitution"
	/replace="ctc"
repeat_region	3217. .3265
	/note="L2 repeat: matches 2699. .2748 of consensus"
variation	3612. .3614
	/note="clone RPL-222E13; cga in this entry; substitution"
	/replace="caa"
misc_feature	3620. .3977
	/note="match: STS: Em:G49294"
variation	3724. .3726
	/note="clone RPL-222E13; tct in this entry; substitution"
	/replace="tat"
variation	3740. .3742
	/note="clone RPL-222E13; gag in this entry; substitution"
	/replace="ggg"
variation	3867. .3869
	/note="clone RPL-222E13; caa in this entry; substitution"
	/replace="cga"
repeat_region	4084. .4394
	/note="AluX repeat: matches 1. .311 of consensus"
variation	4382. .4383
	/note="clone RPL-222E13; ag in this entry; deletion"
	/replace="aaag"
repeat_region	4414. .4720
	/note="AluX repeat: matches 5. .311 of consensus"
variation	4461. .4463
	/note="clone RPL-222E13; gcg in this entry; substitution"
	/replace="gtg"
variation	4470. .4472
	/note="clone RPL-222E13; ttg in this entry; substitution"
	/replace="tcg"
variation	4564. .4566
	/note="clone RPL-222E13; cac in this entry; substitution"
	/replace="cgc"
repeat_region	4723. .4800
	/note="39 copies 2 mer aa 71 conserved"
repeat_region	4947. .5228
	/note="AluX repeat: matches 1. .303 of consensus"
repeat_region	5413. .5602
	/note="AluX repeat: matches 121. .299 of consensus"
variation	5418. .5420
	/note="clone RPL-222E13; ttt in this entry; insertion"
	/replace="ttt"
variation	5598. .5600
	/note="clone RPL-222E13; tga in this entry; substitution"
	/replace="tta"
repeat_region	5603. .5903
	/note="AluX repeat: matches 1. .301 of consensus"
variation	5622. .5623
	/note="clone RPL-222E13; tg in this entry; deletion"
	/replace="ttg"
variation	5641. .5643
	/note="clone RPL-222E13; gcc in this entry; substitution"
	/replace="gtc"
variation	5899. .5901
	/note="clone RP5-919B11; cgg in this entry; substitution"
	/replace="ctg"
repeat_region	5904. .6021
	/note="AluX repeat: matches 1. .121 of consensus"
variation	6088. .6090
	/note="clone RPL-222E13; gcc in this entry; substitution"
	/replace="gtc"
repeat_region	6327. .6636
	/note="AluX repeat: matches 1. .312 of consensus"

```

repeat_region 6662..6899
/note="Alusq repeat: matches 1..301 of consensus"
variation 6970..6972
/note="clone RP1-222E13: aaa in this entry; substitution"
/note="replace"aga"
variation 6987..6989
/note="clone RP1-222E13: att in this entry; substitution"
/note="replace"act"
7045..7109
/note="MLT1H repeat: matches 460..526 of consensus"
/note="complement(7095..7471)"
misc_feature 7142..7161
/note="match: GSS: Em:AQ214568 Em:AQ214591"
repeat_region 7162..7182
/note="FLAM_C repeat: matches 122..142 of consensus"
repeat_region 7187..7482
/note="Alusq repeat: matches 1..311 of consensus"
repeat_region 7493..7616
/note="FLAM_C repeat: matches 1..122 of consensus"
repeat_region 7684..7798
/note="MLT1F repeat: matches 234..354 of consensus"
repeat_region 7823..7885
/note="LTR26 repeat: matches 512..579 of consensus"
repeat_region 7887..8156
/note="Alusq repeat: matches 1..276 of consensus"
misc_feature 8157..8166
/note="match: GSS: Em:AQ394776"
repeat_region 8353..8513
/note="LTR8 repeat: matches 324..478 of consensus"
variation 8509..8511

```

```

Query Match 2.5%: Score 19; DB 9; Length 36339;
Best Local Similarity 100.0%; Pred.No.33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 230 AAGGCCAGGAGCTGACA 248
Db 9869 AAGGCCAGGAGCTGACA 9887
|||||

```

```

RESULT 49
SPAPB1A10
LOCUS SPAPB1A10 36484 bp DNA linear PLN 11-JAN-2001
DEFINITION S.pombe chromosome I BAC pB1A10.
ACCESSION AL512562
VERSION AL512562.1 GI:12188964
KEYWORDS "ras"-related protein rab-7; agglutinin-like; alpha enolase; anaphase
spindle; ATP dependent RNA helicase; cwpl; cytokinesis;
d-arabinono-1,4-lactone oxidase; geranylgeranyl transferase;
glutamyl-trna synthetase; nuclear export factor.
fission yeast.
SOURCE Schizosaccharomyces pombe
ORGANISM Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (bases 1 to 36484)
O'Neill,S., Harris,D., Wood,V., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (06-JAN-2001) European Schizosaccharomycetes genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S.pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was
sequenced by the Sanger Centre. The sequencing of the S. pombe
genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in Pombase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: it is possible that for

```

any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPBC22H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. BAC pB1A10 is overlapped at the 5' end by pl p11E10, EMBL entry SPAP11E10, accession number AL512493, and at the 3' end by cosmid c140, EMBL entry SPAC140, accession number AL163191.

FEATURES

```

Location/Qualifiers
1..36484
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="1"
/map="1L"
/clone="BAC pB1A10"
/complement(1..1672)
/gene="SPAPB1A10.01c"
/note="SPAP11E10.02c"
/complement(1..1672)
/partial
/gene="SPAPB1A10.01c"
/note="SPAPB1A10.01c, SIMILARITY: Saccharomyces
diastaticus, AAC49609, glucanase, (1367 aa), fasta
scores: opt: 516, E(1.1e-18, (29.4% identity in 582 aa)"
/codon_start=1
/label="SPAPB1A10.01c"
/product="hypothetical serine/threonine repeat containing
protein: agglutinin-like"
/protein_id="CAC21474.1"
/db_xref="GI:12188965"
/translation="MSIALAFFILVLGFWSAPSLADDTFNVRSVGLISSNLESC
QSPLEVGYNIYSTASEILTDKAYITIGVSSNSIQDLIDVGNSSNNAASNN
PTSTVTVRVQVTVETVLSQCAQTFVDSISSTSSVINSPTGTVAVSQIITLS
MSPSTPVSFVSASVSKVASVSVSEPSDSSSTNTVLTIVNSPVSSELT
SVLTSTESAVTSSVDIAATASSTASSTVSTSEATVSTFSTDPATPSTLSPASSS
SYLVETSTLTDSVFTTATSDSSVITYTLINVTSSSTTNLPSSSSSLVTIGESS
FPSSLLSLTQSTVSTVSTSSSTQDLTASPISSSVSVSPSPSSILTNSGSIKS
GDRQIVTFSVQTHQSVETLTVITLTITLTITLTDYSHFTLTITTPSPNSISYT
NMTFIPSSSIKSVISVTPVSAENVTSEAFSTSSLVVIPPVNSSLVSTSTSTFTKF
SSLSQSLSTENTTSASSSLTNAKSSLTSTPTTITPTTSSSVSLQTS"
1..122
/misc_feature
/note="nominal overlap with p11E10 S. pombe chromosome 1"
3465..4475
/gene="SPAPB1A10.02"
3465..4475
/gene="SPAPB1A10.02"
/note="SPAPB1A10.02, len:337, SIMILARITY: Gallus gallus,
ENOAC.CHICK, alpha enolase, (433 aa), fasta scores: opt:
141, E(1.0.18, (27.9% identity in 201 aa)"
/codon_start=1
/label="SPAPB1A10.02"
/product="putative alpha enolase"
/protein_id="CAC21475.1"
/db_xref="GI:12188966"
/translation="MCNRQPKAVDLPNYSCHPLLFSQSVERTNPTACDDVCKRIE
SEKYNDFLESLEKKGGRDSDIADEVLDATGEIIVNCHGLEALTKDDIWPFTNNL
EISASNGYKELLOSSIGNPGEKAVSPVHIEDFQSPQIYKFNLSLDEWVSCYFADE
VPLASLVEVNCNETIPQSCVKAVKPKINDKTRKVDASALEKSCLLSDSGTLTDQGL
DTIKHKSIEQNEILHVITDLSPPRNPLLSPPKPLRRSFKVRNSNSTRNE
ISLTSMTSPRNLSHFNLFQPLSQQTSTFGSGSTQNPSSSTCKKAFQCFQISESK
KC"
join(4894..4912,4965..5236,5608..5919)
/gene="SPAPB1A10.03"
join(4894..4912,4965..5236,5608..5919)
/gene="SPAPB1A10.03"

```



```

bronchiseptica), LPS1 (cosmid name), .01 (first CDS), c
(complementary strand)
the more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons
supplemented by a specially developed Hidden Markov Model. CAUTION:
We may not have predicted the correct initiation codon. Where
possible we choose an initiation codon (atg, gtg or ttg) which is
preceded by an upstream ribosome binding site sequence (optimally
5-13bp before the initiation codon). If this cannot be identified
we choose the most upstream initiation codon.

FEATURES
    source
        1..41642 Location/Qualifiers
            /organism="Bordetella bronchiseptica"
            /strain="CN7635E"
            /db_xref="taxon:518"
            /clone="cosmid BbLPS1"
            1..827
            /gene="BbLPS1.01"
            <1..827
            /gene="BbLPS1.01"
            /note="BbLPS1.01, probable formyl transferase, partial
            CDS, len: >274 aa; similar to TR:Q50378 (EMBL:U010425)
            ferric exochelin biosynthesis gene fxbA from Mycobacterium
            smegmatis (360 aa) fasta scores: opt: 325 z-score: 469.7
            E(): 6.6e-19, 31.8% identity in 170 aa overlap and to the
            C-terminus of a hypothetical protein within Salmonella
            typhimurium putative 4-aminocarabiose lipid A modification
            operon TR:Q52325 (EMBL:AF036677) orf3 (660 aa), fasta
            scores: opt: 332 z-score: 379.2 E(): 7.1e-14, 27.9%
            identity in 258 aa overlap. Also similar to many Met-tRNA
            formyltransferases, e.g. FMT_ECOLI methionyl-tRNA
            formyltransferase (314 aa), fasta scores: opt: 335
            z-score: 363.0 E(): 5.8e-13, 29.6% identity in 226 aa
            overlap. C contains Pfam match to entry PF00551
            formyl_transf, Formyl transferase"
            /codon_start=3
            /transl_table=11
            /product="putative formyl transferase"
            /protein_id="CAA07640.1"
            /db_xref="GI:3451484"
            /db_xref="SPTREMBL:O87974"
            /translation="ISYSDKPVAISRQADVAQWVAHAIPAIEYAGVKSVEHFRNP
            ALCVVAGWTHMLPSRLDFRIGAVGLHSLRLPGAGPLNWAILAGPRTAVTLP
            AIGDGVDDGQPVYVQGEAIDVGPNDYIGELHARCNAASVILVERCISGIIIDGLSLPAVQD
            GVASYGQSPRQGVIDMAAPAAAGIARIVRAVSRYPGASTWLDGKLVIRWAGEYRA
            AAVYGAGQITLCPGGQVYVATCKGAIEIEEAEDUGGQGVMMVALRRSHRFRSOS"
            135..440
            /gene="BbLPS1.01"
            /note="Pfam match to entry PF00551 formyl_transf, Formyl
            transferase"
            824..1612
            /gene="BbLPS1.02"
            824..1612
            /gene="BbLPS1.02"
            /note="BbLPS1.02, unknown, len: 262 aa"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein BbLPS1.02"
            /protein_id="CAA07641.1"
            /db_xref="GI:3451485"
            /db_xref="SPTREMBL:O87975"
            /translation="MTGTREGAMETDSALKYHSDFTREAYRSLRLTARASYVRSY
            DTFDPERVLNHDLDPSVHAASALEAEQVSSYFILLHSDFVNLLEVDVSRK
            VESILALGRLHFDGCGYIGISSVAQLEDMLAMEARLLERFVCGRIQVDFSHNPDAW
            MLQQREVRLAGLINTSEFFRQCVGYCSDSNGYWRRLADVLEQAQHSQVLTHPG
            WVDPEMSPRDRIARCAEGRRAATMKKYDEDLACFGDNIGKST"
            1609..2328
            gene

misc_feature
    135..440
    /gene="BbLPS1.01"
    /note="Pfam match to entry PF00551 formyl_transf, Formyl
    transferase"
    824..1612
    /gene="BbLPS1.02"
    824..1612
    /gene="BbLPS1.02"
    /note="BbLPS1.02, unknown, len: 262 aa"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein BbLPS1.02"
    /protein_id="CAA07641.1"
    /db_xref="GI:3451485"
    /db_xref="SPTREMBL:O87975"
    /translation="MTGTREGAMETDSALKYHSDFTREAYRSLRLTARASYVRSY
    DTFDPERVLNHDLDPSVHAASALEAEQVSSYFILLHSDFVNLLEVDVSRK
    VESILALGRLHFDGCGYIGISSVAQLEDMLAMEARLLERFVCGRIQVDFSHNPDAW
    MLQQREVRLAGLINTSEFFRQCVGYCSDSNGYWRRLADVLEQAQHSQVLTHPG
    WVDPEMSPRDRIARCAEGRRAATMKKYDEDLACFGDNIGKST"
    1609..2328
    gene

CDS
    1609..2328
    /gene="BbLPS1.03"
    /note="BbLPS1.03, unknown, len: 239 aa"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein BbLPS1.03"
    /protein_id="CAA07642.1"
    /db_xref="GI:3451486"
    /db_xref="SPTREMBL:O87976"
    /translation="MKFAISIDIDGSPETWRGKVFLSFDIDMAEDFVLLDTLELIER
    AGPATAFWHTQALLERHPFELGHPFNLLSAGSAQAEQVLDALALAPG
    CRSVSHSUTQSTRLLAFADRLGHECNALIPWDAGIPLRPMRWGDTTVRVPHCWE
    DDIALAGAPFLEGDFAFYWDPDGLNVLDFPHPIRVYINTETLERYEASRPVHRDSALP
    AMRHGGQGVRTFLEKILVGAR"
    2313..2317
    RBS
    /gene="BbLPS1.03"
    /note="possible RBS upstream of BbLPS1.04"
    2325..3254
    /gene="BbLPS1.04"
    2325..3254
    /gene="BbLPS1.04"
    /note="BbLPS1.04, probable formyl transferase, len: 309
    aa; similar to e.g. FTDH_RAT formyltetrahydrofolate
    dehydrogenase (formyltransferases domain) (902 aa), fasta
    scores: opt: 199 z-score: 243.2 E(): 2.7e-06, 27.1%
    identity in 225 aa overlap and to e.g. TR:Q50378
    (EMBL:U010425) ferric exochelin biosynthesis gene fxbA from
    Mycobacterium smegmatis (360 aa), fasta scores: opt: 395
    z-score: 264.3 E(): 1.8e-07, 29.4% identity in 286 aa
    overlap. Also similar to the C-terminus of a hypothetical
    protein within Salmonella typhimurium putative
    4-aminocarabiose lipid A modification operon TR:Q52325
    (EMBL:AF036677) orf3 (660 aa), fasta scores: opt: 335
    z-score: 264.3 E(): 1.8e-07, 29.4% identity in 286 aa
    overlap. . Contains Pfam match to entry PF00551
    formyl_transf, Formyl transferase"
    /codon_start=1
    /transl_table=11
    /product="putative formyl transferase"
    /protein_id="CAA07643.1"
    /db_xref="GI:3451487"
    /db_xref="SPTREMBL:O87977"
    /translation="MKFVAIGRHEILLRTIEMLVAGHELAVGLTAAPAEYRADVDD
    FRSLAARKGVPCHVSNRYGPEMVAALAGAEIAYSVNFTYIGHAAIDSPPRGLINA
    HGGDLPRYKGNACQANALIQGPAIGLCVHYVADELSDGVIKAMLDVDHHTTGT
    VAWMGCATPPLVAAALERLRHDPYSVLYRQDDGRRPALRCYPROPSDGRIDMTKPAI
    DVVRHNASGHPYAGAFFIFESNAYRVDEAQWIDGGAYLAVPQVAGRNPDGSDVAC
    GGGLGRLARIADSGAGHWDPVNHIKSMRRLA"
    2514..2855
    /gene="BbLPS1.04"
    /note="Pfam match to entry PF00551 formyl_transf, Formyl
    transferase, score 68.90, E-value 1.1e-18"
    3264..3267
    /note="possible RBS upstream of BbLPS1.05"
    3277..4101
    /gene="BbLPS1.05"
    3277..4101
    /gene="BbLPS1.05"
    /note="BbLPS1.05, probable formyl transferase, len: 274
    aa; some similarity to many Met-tRNA formyltransferases,
    e.g. FMT_BACSU methionyl-tRNA formyltransferase (317 aa),
    fasta scores: opt: 268 z-score: 333.9 E(): 2.4e-11, 26.0%
    identity in 181 aa overlap. Also similar to the C-terminus
    of a hypothetical protein within Salmonella typhimurium
    putative 4-aminocarabiose lipid A modification operon
    TR:Q52325 (EMBL:AF036677) orf3 (660 aa), fasta scores:
    opt: 337 z-score: 411.2 E(): 1.2e-15, 30.4% identity in
    207 aa overlap. . Contains Pfam match to entry PF00551
    formyl_transf, Formyl transferase"
    /codon_start=1
    /transl_table=11
    /product="putative formyl transferase"

```

```

/protein_id="CAA07644.1"
/db_xref="GI:3451488"
/db_xref="SPTREMBL:O87978"
/translation="MTTOVSKPRLAFVDNVCAAELWPAIGTLAPVCALVAAGIRPOY
IFSLADAEKMSCLLIOPRNDPDTYAGFVQGLAQCTAPRILLVHSYKSLRPDVLISY
DYDALNIRHALLPRKRGPNVQWALIHDEAEATGVLYLDDGLDTGDIVAQERIGTSD
ADWTWLSKRLQATGRILNNMAALIDRLPRTPQOAKRETKNSRUTAESPRLDYDG
MSURELFNHWRAQVPLQCGGVYIAGGERLHVRHYVALRDIPALRRRLCAWLEVR"
3541..3798
/misc_feature
  gene="BbLPS1.05"
  note="Pfam match to entry PF00551 formyl_transf, Formyl
  transferase, score 82.00, E-value 1.9e-22"
RBS
  gene="BbLPS1.05"
  note="possible RBS upstream of BbLPS1.06"
Query Match      2.5%; Score 19; DB 1; Length 41642;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 287 AACACGACATTTTCGGT 305
|||||
Db 31960 AACACGACATTTTCGGT 31942

RESULT 51
LOCUS      U88974          43075 bp    DNA    linear    PHG 25-NOV-1998
DEFINITION Streptococcus thermophilus temperate bacteriophage O1205, complete
            genome.
ACCESSION  U88974
VERSION    U88974.1
KEYWORDS   GI:2444080
SOURCE     Streptococcus thermophilus temperate bacteriophage O1205.
            Streptococcus thermophilus temperate bacteriophage O1205
            viruses: dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
REFERENCE  1. (bases 1 to 43075)
            Stanley,E., Fitzgerald,G.F., Le Marrec,C., Fayard,B. and van
            Sinderen,B.
            Sequence analysis and characterization of phi O1205, a temperate
            bacteriophage infecting Streptococcus thermophilus CNR21205
            Microbiology 143 (Pt 11), 3417-3429 (1997)
            98048466
REFERENCE  2. (bases 1 to 43075)
            Stanley,E., Fitzgerald,G.F., Le Marrec,C., Fayard,B. and van
            Sinderen,B.
            Direct Submission
            Submitted (10-FEB-1997) Department of Microbiology, University
            College, Cork, College Road, Cork, Ireland
            Location/Qualifiers
FEATURES             Source
     1. 43075
     /organism="Streptococcus thermophilus temperate
     bacteriophage O1205"
     /strain="CNR21205"
     /db_xref="taxon:85154"
     complement(23..1102)
     /note="putative integrase"
     /codon_start=1
     /transl_table=11
     /evidence=not_experimental
     /product="ORF1"
     /protein_id="AAC79517.1"
     /db_xref="GI:2444081"
     /translation="MASYKRKRSNGWEVRINYDSTGKRPKTKGFRKTSKAAKAAE
     MELQDNVVDIEDILYDFKQWCGVYKPTVSKITYRAYINSQRRIELFFGDKRLK
     SITATEQRVLNSTAKTHAQTVERNVHVKACIEMAVHEGYIKRNFCKFAIKRNT
     GROLETVEEYERLTIYETSKHPEYASYAALYIISKGTIRFAECLGTVDDIKRDT
     GMLSVNKTWYKNNGTGMPKTKSSIREIPEDDFINFIDQLPPTDGGRLPLSLNNA
     VNKTLRIVGVRVSHRLTYASYLIANDIDLSISQVVLGHENLNTITLEVYAHQLE
     QKSRNDEKIKQWTKCONALKPHG"
     complement(1216..1686)
     /codon_start=1
     /transl_table=11

```

```

/product="ORF2"
/protein_id="AAC79518.1"
/db_xref="GI:2444082"
/translation="MKMDQVRNPYTLKAFETTFGCTFGGLILGIFFPFWLLMPVGVA
LIFALFWIKHPTKQEQQLKERAKAKEEFRKSRNEFKKAMAEARAKVCPYCKST
DVEFVQQRKFSJCKAAAGTITMGVGAGLAFAGKKGKKEWHCKNCGAVFTTK"
complement(1706..2092)
/codon_start=1
/transl_table=11
/product="ORF3"
/protein_id="AAC79519.1"
/db_xref="GI:2444083"
/translation="MKIDELLKEYKISLFIIPAEELWERSGYFPDLRIYVNGSLSKQ
EREKVLHIEIGHVHDPRKRLQLQYENQADREMIRELLVYLKSTDIYDFNWRFA
AQYNISTTWGEAMIQDEFKIQQSVI"
complement(2082..2486)
/note="putative repressor"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ORF4"
/protein_id="AAC79520.1"
/db_xref="GI:2444084"
/translation="MKGDVIMKENTAQBLAQIMKERNLRQVDILRQSEKFKQELGINKL
GKSALSIYSGKSIPTQDKLVLLSKTLGVSQAWLMGYDITNESTOVKRESIDLSNLRK
VVMFDKPLSDDDVKKIIEQILRLSIEVTGDED"
2635..2874
/codon_start=1
/transl_table=11
/product="ORF5"
/protein_id="AAC79521.1"
/db_xref="GI:2444085"
/translation="MIKYDYSKINGKIVEFGTKKKFAQSMNLSEKISAKTNKRC
WQSEISKACELLQIPSDIENSYFFRYKVQDIEHN"
2888..3169
/codon_start=1
/transl_table=11
/product="ORF6"
/protein_id="AAC79522.1"
/db_xref="GI:2444086"
/translation="MTDPFKPLADRFNLLITTVIADKSKAFIDITLPLILTAKQCOE
MLGIGNYTEFLRITNLDGFKIDKRGSOVRYPRDQVROWFNHQQDIA"
3230..3358
/codon_start=1
/transl_table=11
/product="ORF7"
/protein_id="AAC79523.1"
/db_xref="GI:2444087"
/translation="MKNLFWILSKDEKQKPVWTPYEVNERYEERHKREQLLMK"
3616..4089
/codon_start=1
/transl_table=11
/product="ORF8"
/protein_id="AAC79524.1"
/db_xref="GI:2444088"
/translation="MATLYELTDQLLEIYNNMDVDETKLDTLEAIDWTIDYENKVEGY
VKVKSLEADIEARKNEKRLDGLNKSDQSKIDKLTALAVSMAETGQTRVDTTLFKV
GFRKSEAVVNEEKLPEYQIATVYKPKTKLKLKSGKHIEGATLEERNLNR"
4086..4787
/note="contains a putative ATP/GTP binding site between
amino acids 21-28"
/codon_start=1
/transl_table=11
/product="ORF9"
/protein_id="AAC79525.1"
/db_xref="GI:2444089"
/translation="MKLIIRANDQRKNRILLYGKAGLCKTSLIKNMPGKTLVLSLD
NSSVLATGENDVIDDFREHTEFEFTQADNLKNYENLVIDNISFSDSMFIE
QGRKSGNSNELQHSQWNTYFLRVLTIVYKPINIYVTAMEDTHELNETGQLTQ
YVQIRASVLNQLLGDYVGVIVNAKTGARGLLEGSEGTYAKNRLNRTACKIED
LKFEGDLDTGKELPE"
4762..6093
/note="putative helicase"

```

```

/codon_start=1
/transl_table=11
/evidence=not experimental
/product="ORF10"
/protein_id="AAC79526.1"
/db_xref="GI:2444096"
/translation="MELRNYQSDLVNDIKQSILRCKRMVQSPRSGKTVYMAHIAK
GATDKLIFESHRKEINEOVNTFRNGVDMNLVTIDSVTKIARNLDRIOEPSIIL
IDEAHVKAQTYKIIEYNSIVLMFTGTGPARLGGSCFDIADIVLGSVKKWLODN
GNAPEKYAPSLIDITNLKRGGEFTKKSVDUTMKRVLYGDIYRHKVAKGKQALV
YTHSVEASESVNTNEQGYTSIAISGKTPEVRDRAQAFROGKLTIMYNCELFTEG
IDLNVDCIMLRPTQSLVLOFAMRALNPRBKTAIIDHVGNDVHGLPNADREW
SLKGVNKKLKLCEPTRYCDECIATFNSAERIKCMCHENAPTRKEEVIREIEL
ERRREVASKEVTFVSDQCQVELKREFAKHGKYGKPGVYVYQKKNWR"
6100..6555
/feature="function unknown"
/codon_start=1
/transl_table=11
/product="ORF11"
/protein_id="AAC79527.1"
/db_xref="GI:2444091"
/translation="MFTIDYSOAKEFGSIKDGTYEVTIDIAKQDATOGGADYLDIRFR
IRKDFQEQFQNNIIFIRIFAKKEDGKYPVASIMNIAKAGIPDGTKESSIEDLYNOLE
GKALVTVKNEKSEWQGTYENLVNKRLEVTIDIPLEVNTSEISIDLPF"
6564..7373
/codon_start=1
/transl_table=11
/product="ORF12"
/protein_id="AAC79528.1"
/db_xref="GI:2444092"
/translation="MVDYAINVQRMGYSVIPISKNGKTPILISFADKPPMTENDIRRVW
RNDPANIALKTDFVFDVDMHGDVDGLNLRNWEHARLIPPTLQAITPSGGRHLYL
KDPNHPISONGITEGVDIKAHVNNVILPPSNKGYVEMDTVHSPKDGSTIEAPL
ALIKVLOAKMKEPLSVSESPADCVSRSTIKLFESILLQFGDKGGRNALAEFVGG
ILLRGVDPEIYHLAKMANNTQPLGDKTEFRTKSLMDKEIRRIGDND"
7360..8874
/feature="putative primase"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="ORF13"
/protein_id="AAC79529.1"
/db_xref="GI:2444093"
/translation="MTTIDFYREQFASSLSPGKPSREGINKKLKAVRNDWPKFK
KEENPKKPELAPLAVAKGLNKYTHVITLNGKVAIYDPERGYOKDYRYAVGLIY
VLEPTFNKTRNLVFLSLNSREYENNMVDFEYRDVRRFILLKGLYDKRKK
LLSDYKFINFTTETIEVENAPKPTINGMDVDSWLDLMSGSEIYELLQVLAASL
NGNHSYRKSINLVGNDGDKCTEOLISNLVGLKNVAPKINQFSERFGLAITEGTV
LLCDVQAGIYVDSNSPNSVWGEPVSEIKKGNPILAQFKATVIOSTNAMPVFNK
SNGTRRIIIPFKTEGINDNMAIKDDYINRKEVILEYVILKAIILDFDKFSEPKAT
QERCEPKEENNTVYKLENYLSDVSTRIPVRFMDVYRSCHEGNTIPKKSINPEK
FLAQLNPVQWTKDKQKPLDFPNFKDKPVYVHDFNFWNDENEAKKAAVVVMVQ"
9128..9559
/codon_start=1
/transl_table=11
/product="ORF14"
/protein_id="AAC79530.1"
/db_xref="GI:2444094"
/translation="MEIQYLEINOEHEPENISNFKDFSEATVIDVQCNAIPVHFE
KVGEDYTDDEYGVKVVAFIKYEDNKEATEPKKWLKEPFEDKHDKETLFLDIDDPDN
WATKGDVYVDFNCLFQQAASMYATDKQIVSIYKEMLQNH"
9601..9861
/codon_start=1
/transl_table=11
/product="ORF15"
/protein_id="AAC79531.1"
/db_xref="GI:2444095"
/translation="MVFRINVGKVRMADGWEDTGAPKGCDFLFGFRPDGQIFFLEVK
NEKRLRDDOKKFEAMKRGALVGVARSVKQAMDIVNEKK"
9845..10081
/codon_start=1
/transl_table=11
/product="ORF16"

```

```

/protein_id="AAC79532.1"
/db_xref="GI:2444096"
/translation="MRKNVKKLVGIYAPGHYEHISVLEETQEFSEFWFDHNEHMEFI
SSKLGISTNKLNRILITLEQLPDELLARMGELCK"
10072..10245
/codon_start=1
/transl_table=11
/product="ORF17"

CDS

Query Match      2.5%   Score 19; DB 7; Length 43075;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 AAGCTGAGAAATTCCTCAA 258
|||||
Db 13944 AAGCTGAGAAATTCCTCAA 13962

RESULT 52
AC102174
LOCUS
DEFINITION Mus musculus clone RP23-472123, LOW-PASS SEQUENCE SAMPLING.
AC102174
ACCESSION
VERSION AC102174.1 GI:17061260
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 58747)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-472123
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 58747)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Bouckigalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepe,I., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K.,
Lamazaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylot,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18167
Center clone name: 472_I_23
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into

```

* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
782 881: contig of 781 bp in length
882 881: gap of 100 bp
882 1579: contig of 698 bp in length
1580 1679: gap of 100 bp
1680 2392: contig of 713 bp in length
2393 2492: gap of 100 bp
2493 3207: contig of 715 bp in length
3208 3307: gap of 100 bp
3308 4017: contig of 710 bp in length
4018 4117: gap of 100 bp
4118 4820: contig of 703 bp in length
4821 4920: gap of 100 bp
4921 5632: contig of 712 bp in length
5633 5732: gap of 100 bp
5733 6463: contig of 731 bp in length
6464 6563: gap of 100 bp
6564 7262: contig of 699 bp in length
7263 7362: gap of 100 bp
7363 8065: contig of 703 bp in length
8066 8165: gap of 100 bp
8166 8889: contig of 704 bp in length
8890 8969: gap of 100 bp
8970 9678: contig of 709 bp in length
9679 9778: gap of 100 bp
9779 10492: contig of 714 bp in length
10493 10592: gap of 100 bp
10593 11308: contig of 716 bp in length
11309 11408: gap of 100 bp
11409 12086: contig of 678 bp in length
12087 12186: gap of 100 bp
12187 12850: contig of 694 bp in length
12881 12980: gap of 100 bp
12981 13693: contig of 713 bp in length
13694 13793: gap of 100 bp
13794 14569: contig of 776 bp in length
14570 14669: gap of 100 bp
14570 15359: contig of 690 bp in length
15360 15459: gap of 100 bp
15460 16152: contig of 693 bp in length
16153 16252: gap of 100 bp
16253 16950: contig of 698 bp in length
16951 17050: gap of 100 bp
17051 17760: contig of 710 bp in length
17761 17860: gap of 100 bp
17861 18578: contig of 718 bp in length
18579 18678: gap of 100 bp
18679 19394: contig of 716 bp in length
19395 19494: gap of 100 bp
19495 20189: contig of 695 bp in length
20190 20289: gap of 100 bp
20290 20991: contig of 702 bp in length
20992 21091: gap of 100 bp
21092 21799: contig of 708 bp in length
21800 21899: gap of 100 bp
21900 22595: contig of 696 bp in length
22596 22695: gap of 100 bp
22696 23408: contig of 713 bp in length
23409 23508: gap of 100 bp
23509 24199: contig of 691 bp in length
24200 24299: gap of 100 bp
24300 24995: contig of 696 bp in length
24996 25095: gap of 100 bp
25096 25803: contig of 708 bp in length
25804 25903: gap of 100 bp

25904 26619: contig of 716 bp in length
26620 26719: gap of 100 bp
26720 27444: contig of 725 bp in length
27445 27544: gap of 100 bp
27545 28235: contig of 691 bp in length
28236 28335: gap of 100 bp
28336 29035: contig of 704 bp in length
29040 29139: gap of 100 bp
29140 29829: contig of 690 bp in length
29830 29929: gap of 100 bp
29930 30637: contig of 708 bp in length
30638 30737: gap of 100 bp
30738 31439: contig of 702 bp in length
31440 31539: gap of 100 bp
31540 32234: contig of 695 bp in length
32235 32334: gap of 100 bp
32336 33039: contig of 705 bp in length
33040 33139: gap of 100 bp
33140 33851: contig of 712 bp in length
33852 33951: gap of 100 bp
33952 34669: contig of 718 bp in length
34670 34769: gap of 100 bp
34770 35480: contig of 711 bp in length
35481 35580: gap of 100 bp
35581 36292: contig of 712 bp in length
36293 36392: gap of 100 bp
36393 37104: contig of 712 bp in length
37105 37204: gap of 100 bp
37205 37912: contig of 708 bp in length
37913 38012: gap of 100 bp
38013 38714: contig of 702 bp in length
38715 38814: gap of 100 bp
38815 39508: contig of 694 bp in length
39509 39608: gap of 100 bp
39609 40303: contig of 695 bp in length
40304 40403: gap of 100 bp
40404 41103: contig of 700 bp in length
41104 41203: gap of 100 bp
41204 41914: contig of 711 bp in length
41915 42014: gap of 100 bp
42015 42712: contig of 698 bp in length
42713 42812: gap of 100 bp
42813 43520: contig of 708 bp in length
43521 43620: gap of 100 bp
43621 44316: contig of 696 bp in length
44317 44416: gap of 100 bp
44417 45124: contig of 708 bp in length
45125 45224: gap of 100 bp
45225 45918: contig of 694 bp in length
45919 46018: gap of 100 bp
46019 46716: contig of 698 bp in length
46717 46816: gap of 100 bp
46817 47502: contig of 686 bp in length
47503 47602: gap of 100 bp
47603 48307: contig of 705 bp in length
48308 48407: gap of 100 bp
48408 49118: contig of 711 bp in length
49119 49218: gap of 100 bp
49219 49930: contig of 712 bp in length
49931 50030: gap of 100 bp
50031 50737: contig of 707 bp in length
50738 50837: gap of 100 bp
50838 51535: contig of 698 bp in length
51536 51635: gap of 100 bp
51636 52345: contig of 710 bp in length
52346 52445: gap of 100 bp
52446 53146: contig of 701 bp in length
53147 53246: gap of 100 bp
53247 53948: contig of 702 bp in length
53949 54048: gap of 100 bp
54049 54763: contig of 715 bp in length
54764 54863: gap of 100 bp
54864 55559: contig of 696 bp in length

```

* 55560 55659: gap of 100 bp
Query Match 2.5%: Score 19; DB 2; Length 56747;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 GACTTGACAGAGTTGACAA 147
|||||
Db 29608 GACTTGACAGAGTTGACAA 29626

RESULT 53
AL392047/c
LOCUS
DEFINITION Human DNA sequence from clone RP11-189W8 on chromosome 10, complete
sequence.
ACCESSION AL392047
VERSION AL392047.11 GI:17426975
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 59314)
Direct Submission
Submitted (07-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Dec 8, 2001 this sequence version replaced gi:15020939.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/chr10
RP11-189W8 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-189W8. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP5-933E2 is at 57315 in this sequence.
The true right end of clone RP11-675023 is at 2000 in this
sequence.

FEATURES
Location/Qualifiers
1..59314
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-189W8"
/clone.lib="RPCI-11.1"
BASE COUNT 18957 a 10703 c 11134 g 18520 t
ORIGIN
```

```

Query Match 2.5%: Score 19; DB 9; Length 59314;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 ACACAGCCAAATCATCAAC 583
|||||
Db 41192 ACACAGCCAAATCATCAAC 41174

RESULT 54
AL107830/c
LOCUS
DEFINITION Mus musculus clone RP23-230L20, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC107830
VERSION AC107830.1 GI:18308533
KEYWORDS HTG: HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 62832)
Unpublished
2 (bases 1 to 62832)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginder, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Rise, C., Rogov, P., Roman, J.,
Retta, K., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20580
Center clone name: 230_L-20
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
```



```
* 680 779: contig of 679 bp in length
* 780 779: gap of 100 bp
* 1489 1588: contig of 709 bp in length
* 1589 2358: contig of 670 bp in length
* 2359 3178: contig of 720 bp in length
* 3179 3850: contig of 672 bp in length
* 3851 3950: gap of 100 bp
* 3951 4648: contig of 698 bp in length
* 4649 4748: gap of 100 bp
* 4749 5467: contig of 719 bp in length
* 5468 5567: gap of 100 bp
* 5568 6272: contig of 705 bp in length
* 6273 6372: gap of 100 bp
* 6373 7072: contig of 700 bp in length
* 7073 7172: gap of 100 bp
* 7173 7912: contig of 740 bp in length
* 7913 8012: gap of 100 bp
* 8013 8333: contig of 721 bp in length
* 8334 8833: gap of 100 bp
* 8834 9657: contig of 724 bp in length
* 9658 10374: contig of 717 bp in length
* 10375 10474: gap of 100 bp
* 10475 11180: contig of 706 bp in length
* 11181 11280: gap of 100 bp
* 11281 11985: contig of 705 bp in length
* 11986 12085: gap of 100 bp
* 12086 12795: contig of 710 bp in length
* 12796 12895: gap of 100 bp
* 12896 13598: contig of 703 bp in length
* 13599 13698: gap of 100 bp
* 13699 14393: contig of 695 bp in length
* 14394 14493: gap of 100 bp
* 14494 15131: contig of 638 bp in length
* 15132 15231: gap of 100 bp
* 15232 15894: contig of 663 bp in length
* 15895 15994: gap of 100 bp
* 15995 16697: contig of 703 bp in length
* 16698 16797: gap of 100 bp
* 16798 17494: contig of 697 bp in length
* 17495 17594: gap of 100 bp
* 17595 18302: contig of 708 bp in length
* 18303 18402: gap of 100 bp
* 18403 19127: contig of 725 bp in length
* 19128 19227: gap of 100 bp
* 19228 19930: contig of 703 bp in length
* 19931 20030: gap of 100 bp
* 20031 20749: contig of 719 bp in length
* 20750 20849: gap of 100 bp
* 20850 21567: contig of 718 bp in length
* 21568 21667: gap of 100 bp
* 21668 22367: contig of 700 bp in length
* 22368 22467: gap of 100 bp
* 22468 23190: contig of 723 bp in length
* 23191 23290: gap of 100 bp
* 23291 24004: contig of 714 bp in length
* 24005 24104: gap of 100 bp
* 24105 24834: contig of 730 bp in length
* 24835 24934: gap of 100 bp
* 24935 25667: contig of 733 bp in length
* 25668 25767: gap of 100 bp
* 25768 26468: contig of 701 bp in length
* 26469 26568: gap of 100 bp
* 26569 27290: contig of 722 bp in length
* 27291 27390: gap of 100 bp
* 27391 28089: contig of 699 bp in length
* 28090 28189: gap of 100 bp
* 28190 28876: contig of 687 bp in length
* 28877 28976: gap of 100 bp
* 28977 29681: contig of 705 bp in length
```

```
* 29682 29781: gap of 100 bp
* 29782 30478: contig of 697 bp in length
* 30479 30578: gap of 100 bp
* 30579 31264: contig of 686 bp in length
* 31265 31364: gap of 100 bp
* 31365 32071: contig of 707 bp in length
* 32072 32171: gap of 100 bp
* 32172 32878: contig of 707 bp in length
* 32879 32978: gap of 100 bp
* 32979 33695: contig of 717 bp in length
* 33696 33795: gap of 100 bp
* 33796 34515: contig of 720 bp in length
* 34516 34615: gap of 100 bp
* 34616 35324: contig of 709 bp in length
* 35325 35424: gap of 100 bp
* 35425 36153: contig of 729 bp in length
* 36154 36253: gap of 100 bp
* 36254 36978: contig of 725 bp in length
* 36979 37078: gap of 100 bp
* 37079 37778: contig of 700 bp in length
* 37779 37878: gap of 100 bp
* 37879 38573: contig of 695 bp in length
* 38574 38673: gap of 100 bp
* 38674 39389: contig of 716 bp in length
* 39390 39489: gap of 100 bp
* 39490 40228: contig of 739 bp in length
* 40229 40328: gap of 100 bp
* 40329 41013: contig of 685 bp in length
* 41014 41113: gap of 100 bp
* 41114 41835: contig of 722 bp in length
* 41836 41935: gap of 100 bp
* 41936 42649: contig of 714 bp in length
* 42650 42749: gap of 100 bp
* 42750 43451: contig of 702 bp in length
* 43452 43551: gap of 100 bp
* 43552 44262: contig of 711 bp in length
* 44263 44362: gap of 100 bp
* 44363 45076: contig of 714 bp in length
* 45077 45176: gap of 100 bp
* 45177 45879: contig of 703 bp in length
* 45880 45979: gap of 100 bp
* 45980 46699: contig of 720 bp in length
* 46700 46799: gap of 100 bp
* 46800 47514: contig of 715 bp in length
* 47515 47614: gap of 100 bp
* 47615 48320: contig of 706 bp in length
* 48321 48420: gap of 100 bp
* 48421 49135: contig of 715 bp in length
* 49136 49235: gap of 100 bp
* 49236 49916: contig of 683 bp in length
* 49917 50016: gap of 100 bp
* 50017 50728: contig of 712 bp in length
* 50729 50828: gap of 100 bp
* 50829 51531: contig of 703 bp in length
* 51532 51631: gap of 100 bp
* 51632 52323: contig of 692 bp in length
* 52324 52423: gap of 100 bp
* 52424 53144: contig of 721 bp in length
* 53145 53244: gap of 100 bp
* 53245 53932: contig of 688 bp in length
* 53933 54032: gap of 100 bp
* 54033 54743: contig of 711 bp in length
* 54744 54843: gap of 100 bp
* 54844 55562: contig of 719 bp in length
* 55563 55662: gap of 100 bp
```

Query Match 2.5%: Score 19; DB 2; Length 62832;
Best Local Similarity 100.0%; Pred.No. 33;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 ACAATACTCTTAAAGGCAT 162
|||||
DB 53511 ACAATACTCTTAAAGGCAT 53493

RESULT 55

AC109187/c

LOCUS

DEFINITION

AC109187

AC109187.1

GI:18482108

HTG: HTGS_PHASE0.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 73648)

Biren, B., Linton, L., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-329N2

Unpublished

2 (bases 1 to 73648)

Biren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kellis, C., Lakocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,

Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,

McSwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,

Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,

Retta, R., Rieback, M., Riley, K., Rise, C., Rogov, P., Roman, J.,

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research,

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20049

Center clone name: 329_N_2

* NOTE: this record contains 91 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be generic and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 654: contig of 654 bp in length

* 655 754: gap of 100 bp

* 755 1486: contig of 732 bp in length

* 1487 1586: gap of 100 bp

* 1587 2317: contig of 731 bp in length

* 2318 2417: gap of 100 bp

* 2418 3145: contig of 728 bp in length

AC109187 73648 bp INA linear HTG 03-FEB-2002
Mus musculus clone RP24-329N2, LOW-PASS SEQUENCE SAMPLING.

AC109187

AC109187.1

GI:18482108

HTG: HTGS_PHASE0.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 73648)

Biren, B., Linton, L., Nusbaum, C. and Lander, E.

Mus musculus, clone RP24-329N2

Unpublished

2 (bases 1 to 73648)

Biren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kellis, C., Lakocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,

Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,

McSwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,

Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,

Retta, R., Rieback, M., Riley, K., Rise, C., Rogov, P., Roman, J.,

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research,

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20049

Center clone name: 329_N_2

* NOTE: this record contains 91 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be generic and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 654: contig of 654 bp in length

* 655 754: gap of 100 bp

* 755 1486: contig of 732 bp in length

* 1487 1586: gap of 100 bp

* 1587 2317: contig of 731 bp in length

* 2318 2417: gap of 100 bp

* 2418 3145: contig of 728 bp in length

* 3146 31559: contig of 697 bp in length

* 31560 32264: contig of 705 bp in length

3146 3245: gap of 100 bp
3246 3947: contig of 702 bp in length
3948 4047: gap of 100 bp
4048 4740: contig of 693 bp in length
4741 4840: gap of 100 bp
4841 5530: contig of 690 bp in length
5531 5630: gap of 100 bp
5631 6340: contig of 710 bp in length
6341 6440: gap of 100 bp
6441 7161: contig of 721 bp in length
7162 7261: gap of 100 bp
7262 7989: contig of 728 bp in length
7990 8089: gap of 100 bp
8090 8829: contig of 740 bp in length
8830 8929: gap of 100 bp
8930 9670: contig of 741 bp in length
9671 9770: gap of 100 bp
9771 10425: contig of 655 bp in length
10426 10525: gap of 100 bp
10526 11249: contig of 724 bp in length
11250 11349: gap of 100 bp
11350 12081: contig of 732 bp in length
12082 12181: gap of 100 bp
12182 12894: contig of 713 bp in length
12895 12994: gap of 100 bp
12995 13689: contig of 695 bp in length
13690 13789: gap of 100 bp
13790 14478: contig of 689 bp in length
14479 14578: gap of 100 bp
14579 15273: contig of 695 bp in length
15274 15373: gap of 100 bp
15374 16089: contig of 716 bp in length
16090 16189: gap of 100 bp
16190 16838: contig of 649 bp in length
16839 16938: gap of 100 bp
16939 17640: contig of 702 bp in length
17641 17740: gap of 100 bp
17741 18455: contig of 715 bp in length
18456 18555: gap of 100 bp
18556 19286: contig of 731 bp in length
19287 19386: gap of 100 bp
19387 20069: contig of 683 bp in length
20070 20169: gap of 100 bp
20170 20900: contig of 731 bp in length
20901 21000: gap of 100 bp
21001 21734: contig of 734 bp in length
21735 21834: gap of 100 bp
21835 22526: contig of 692 bp in length
22527 22626: gap of 100 bp
22627 23317: contig of 691 bp in length
23318 23417: gap of 100 bp
23418 24116: contig of 699 bp in length
24117 24216: gap of 100 bp
24217 24939: contig of 723 bp in length
24940 25039: gap of 100 bp
25040 25769: contig of 730 bp in length
25770 25869: gap of 100 bp
25870 26618: contig of 747 bp in length
26617 26716: gap of 100 bp
26717 27397: contig of 681 bp in length
27398 27497: gap of 100 bp
27498 28243: contig of 746 bp in length
28244 28343: gap of 100 bp
28344 29051: contig of 708 bp in length
29052 29151: gap of 100 bp
29152 29865: contig of 714 bp in length
29866 29965: gap of 100 bp
29966 30662: contig of 697 bp in length
30663 30762: gap of 100 bp
30763 31459: contig of 697 bp in length
31460 31559: gap of 100 bp
31560 32264: contig of 705 bp in length
32265 32364: gap of 100 bp

* 32365 33090: contig of 726 bp in length
 * 33091 33190: gap of 100 bp
 * 33191 33835: contig of 645 bp in length
 * 33836 33935: gap of 100 bp
 * 33936 34649: contig of 714 bp in length
 * 34650 34749: gap of 100 bp
 * 34750 35487: contig of 738 bp in length
 * 35488 35587: gap of 100 bp
 * 35588 36324: contig of 737 bp in length
 * 36325 36424: gap of 100 bp
 * 36425 37108: contig of 684 bp in length
 * 37109 37208: gap of 100 bp
 * 37209 37925: contig of 717 bp in length
 * 37926 38025: gap of 100 bp
 * 38026 38757: contig of 732 bp in length
 * 38758 38857: gap of 100 bp
 * 38858 39577: contig of 720 bp in length
 * 39578 39677: gap of 100 bp
 * 39678 40366: contig of 689 bp in length
 * 40367 40466: gap of 100 bp
 * 40467 41158: contig of 692 bp in length
 * 41159 41258: gap of 100 bp
 * 41259 41971: contig of 713 bp in length
 * 41972 42071: gap of 100 bp
 * 42072 42791: contig of 720 bp in length
 * 42792 42891: gap of 100 bp
 * 42892 43585: contig of 694 bp in length
 * 43586 43685: gap of 100 bp
 * 43686 44417: contig of 732 bp in length
 * 44418 44517: gap of 100 bp
 * 44518 45254: contig of 737 bp in length
 * 45255 45354: gap of 100 bp
 * 45355 46032: contig of 678 bp in length
 * 46033 46132: gap of 100 bp
 * 46133 46853: contig of 721 bp in length
 * 46854 46953: gap of 100 bp
 * 46954 47634: contig of 681 bp in length
 * 47635 47734: gap of 100 bp
 * 47735 48429: contig of 695 bp in length
 * 48430 48529: gap of 100 bp
 * 48530 49225: contig of 696 bp in length
 * 49226 49325: gap of 100 bp
 * 49326 50020: contig of 695 bp in length
 * 50021 50120: gap of 100 bp
 * 50121 50838: contig of 718 bp in length
 * 50839 50938: gap of 100 bp
 * 50939 51581: contig of 643 bp in length
 * 51582 51681: gap of 100 bp
 * 51682 52399: contig of 718 bp in length
 * 52400 52499: gap of 100 bp
 * 52500 53238: contig of 739 bp in length
 * 53239 53338: gap of 100 bp
 * 53339 54064: contig of 726 bp in length
 * 54065 54164: gap of 100 bp
 * 54165 54850: contig of 686 bp in length
 * 54851 54950: gap of 100 bp
 * 54951 55687: contig of 737 bp in length
 * 55688 55787: gap of 100 bp

Query Match 2.5%; Score 19; DB 2; Length 73648;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 CAGGAAGCTGAGAAATTCG 254
 ||||||||||||||||
 Db 56958 CAGGAAGCTGAGAAATTCG 56940

RESULT 56
 AL590432/c 95750 bp DNA linear PRI 16-JAN-2002
 LOCUS Human DNA sequence from clone RP11-141A19 on chromosome 1, complete
 DEFINITION sequence.

ACCESSION AL590432
 VERSION AL590432.9 GI:18250506
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 95760)
 Ellington.A

Direct Submission
 Submitted (16-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Jan 19, 2002 this sequence version replaced gi:16973071.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30).
 An attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
 database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1

RP11-141A19 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-141A19 it may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-141A19 is at 1 in this sequence.
 The true left end of clone RP5-885E17 is at 93761 in this sequence.

FEATURES
 Location/Qualifiers
 1..95760
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-141A19"
 /clone_lib="RPCI-11.1"
 84894..85187
 misc_feature
 /note="Sequence from overlapping clone RP11-450E3
 (AL506474) Assembly confirmed by restriction digest."

BASE COUNT 30572 a 18799 c 17532 g 28857 t
 ORIGIN

source

1..95760
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-141A19"
 /clone_lib="RPCI-11.1"
 84894..85187

misc_feature

/note="Sequence from overlapping clone RP11-450E3
 (AL506474) Assembly confirmed by restriction digest."

BASE COUNT

ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 95760;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 AATCTTGGCAATGACCTAA 89
 ||||||||||||||||

Db 68640 AATCTTGGCAATGACCTAA 68622

RESULT 57
 AP003815/c 106253 bp DNA linear HTG 04-JUL-2001
 LOCUS Oryza sativa chromosome 7 clone OJ1163_G04, *** SEQUENCING IN
 DEFINITION PROGRESS ***, in ordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP003815
AP003815.1 GI:14595155
HTG: HTGS PHASE2
Oriza sativa
Oriza sativa

REFERENCE
AUTHORS
TITLE

Eukaryote: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 106253)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oriza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
clone:OJ1163.G04

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Published Only in Database (2001) In press
2 (bases 1 to 106253)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2 Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@abrr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source

Location/Qualifiers
1..106253
/organism="Oriza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="7"
/clone="OJ1163.G04"
BASE COUNT 30665 a 21666 c 22767 g 31155 t
ORIGIN

Query Match 2.5%; Score 19; DB 2; Length 106253;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 GCAACTGCCAAGCAGAG 232

Db 51962 GCAACTGCCAAGCAGAG 51944

RESULT 58
AP002091/C

LOCUS
DEFINITION

AP002091 106763 bp DNA linear PRI 30-OCT-2001
Homo sapiens genomic DNA, chromosome 6q25.2, clone: C1TD-2058112,
complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AP002091
AP002091.2 GI:15208256
Homo sapiens DNA, clone:lib:CIT Approved Human Genomic Sperm
Library D clone:C1TD-2058112.

ORGANISM

Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 106763)
Shimizu, N. and Asakawa, S.
Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (22-MAY-2000) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
Tel:81-3-3351-2370, Fax:81-3-3351-2370)

COMMENT
FEATURES

On Aug 16, 2001 this sequence version replaced gi:8096474.

Location/Qualifiers
Source

1..106763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6q25.2"
/clone="C1TD-2058112"
148..499
/rpt_family="THEIA"
/evidence=not_experimental
843..1125
/rpt_family="AluSg"
/evidence=not_experimental
complement(2227..2493)
/rpt_family="AluSx"
/evidence=not_experimental
2713..3024
/rpt_family="AluSg"
/evidence=not_experimental
3027..3133
/rpt_family="MLTIG1"
/evidence=not_experimental
3465..3499
/rpt_family="AT-rich"
/evidence=not_experimental
complement(3706..4131)
/rpt_family="MSTB"
/evidence=not_experimental
complement(4260..4558)
/rpt_family="AluY"
/evidence=not_experimental
4847..5046
/rpt_family="MIR"
/evidence=not_experimental
5587..5886
/rpt_family="AluSc"
/evidence=not_experimental
complement(6350..6557)
/rpt_family="L1ME"
/evidence=not_experimental
complement(6640..6750)
/rpt_family="L1ME"
/evidence=not_experimental
6751..7043
/rpt_family="AluSg"
/evidence=not_experimental
complement(7044..7075)
/rpt_family="L1ME"
/evidence=not_experimental
7085..7108
/rpt_family="(TTTTG)n"
/evidence=not_experimental
complement(7109..7388)
/rpt_family="AluY"
/evidence=not_experimental
complement(7389..7827)
/rpt_family="L1ME"
/evidence=not_experimental
8049..8251
/rpt_family="MSTA"
/evidence=not_experimental
8281..8517
/rpt_family="MSTA"
/evidence=not_experimental
8752..8784
/rpt_family="AT-rich"
/evidence=not_experimental
9041..9209
/rpt_family="MIR"
/evidence=not_experimental
9211..9244

```

/rpt_family="(TC)n"
/evidence=not_experimental
repeat_region
9244..9271
/rpt_family="(TA)n"
/evidence=not_experimental
repeat_region
10750..10830
/rpt_family="MIR"
/evidence=not_experimental
repeat_region
11137..11175
/rpt_family="AT-rich"
/evidence=not_experimental
repeat_region
11602..11908
/rpt_family="AluSd1"
/evidence=not_experimental
repeat_region
13292..13577
/rpt_family="AluJo"
/evidence=not_experimental
repeat_region
14159..14311
/rpt_family="MEK5B"
/evidence=not_experimental
repeat_region
15530..15605
/rpt_family="CT-rich"
/evidence=not_experimental
repeat_region
15609..15872
/rpt_family="L2"
/evidence=not_experimental
repeat_region
16304..16537
/rpt_family="MIR"
/evidence=not_experimental
repeat_region
17545..17847
/rpt_family="AluX"
/evidence=not_experimental
repeat_region
18093..18126
/rpt_family="AT-rich"
/evidence=not_experimental
repeat_region
18129..18416
/rpt_family="AluYb8"
/evidence=not_experimental
repeat_region
18457..18506
/rpt_family="AT-rich"
/evidence=not_experimental
repeat_region
18773..18875
/rpt_family="L1PA16"
/evidence=not_experimental
repeat_region
18935..19236
/rpt_family="AluX"
/evidence=not_experimental
repeat_region
19244..19267
/rpt_family="(GGGA)n"
/evidence=not_experimental
repeat_region
19277..19603
/rpt_family="L1PA16"
/evidence=not_experimental
repeat_region
19604..20394
/rpt_family="LTR1"
/evidence=not_experimental
repeat_region
20395..21725
/rpt_family="L1PA16"
/evidence=not_experimental
repeat_region
21969..21992
/rpt_family="(TA)n"
/evidence=not_experimental
repeat_region
22191..22311
/rpt_family="(TA)n"
/evidence=not_experimental
repeat_region
22313..22608
/rpt_family="AluX"
/evidence=not_experimental
repeat_region
23055..23098
/rpt_family="(T)n"
/evidence=not_experimental
repeat_region
23395..23693
/rpt_family="AluSq"

```

```

/evidence=not_experimental
25106..25411
/rpt_family="AluY"
/evidence=not_experimental
repeat_region
25985..26220
/rpt_family="MIR"
/evidence=not_experimental
repeat_region
27269..27291
/rpt_family="AT-rich"
/evidence=not_experimental
repeat_region
28687..29229
/rpt_family="MEK76"
/evidence=not_experimental
repeat_region
29695..30982
/rpt_family="L1MC4"
/evidence=not_experimental
repeat_region
30984..31037
/rpt_family="(TA)n"
/evidence=not_experimental
repeat_region
3115..31235
/rpt_family="L1MC4"
/evidence=not_experimental
repeat_region
31450..31514
/rpt_family="(TCTA)n"
/evidence=not_experimental
repeat_region
31669..31693
/rpt_family="AT-rich"
/evidence=not_experimental
repeat_region
31900..32060
/rpt_family="(TA)n"
/evidence=not_experimental
repeat_region
32061..32188
/rpt_family="(CATATA)n"
/evidence=not_experimental
repeat_region
32189..32210
/rpt_family="AT-rich"

```

Query Match 2.5%; Score 19; DB 9; Length 106763;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

C 3 TGCTTCAGCTTTGGAAACT 21

|||||

Db 77988 TGCTTCAGCTTTGGAAACT 77970

RESULT 59

AC094856/C

LOCUS

DEFINITION

AC094856

AC094856.2

GT:17941639

HTG: HTGS_PHASE1

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 125018)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R., Chen,Z.,

Chondry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

AC094856 125018 bp DNA linear HTG 20-DEC-2001

Rattus norvegicus clone CH230-6B13, *** SEQUENCING IN PROGRESS ***,

69 unordered pieces.

AC094856

AC094856.2

GT:17941639

HTG: HTGS_PHASE1

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 125018)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R., Chen,Z.,

Chondry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, J., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisghe, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S., Ogutu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, R., Thomas, N., Thomas, S., Umanli, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
2 (bases 1 to 125018)
Worley, K.C.
Submitted (15-Sep-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624692.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSOQ
Center clone name: CH230-6B13
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to findPhrapList
Consensus quality: 91161 bases at least Q40
Consensus quality: 103712 bases at least Q30
Consensus quality: 110661 bases at least Q20
Estimated insert size: 90316; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 69 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will be preserved.
1 3519: contig of 3519 bp in length
3520 3619: gap of unknown length
3620 7769: contig of 4150 bp in length
7770 7869: gap of unknown length
7870 10961: contig of 3092 bp in length
10962 11061: gap of unknown length
11062 13257: contig of 2196 bp in length
13258 13357: gap of unknown length
13358 15779: contig of 2422 bp in length
15780 15879: gap of unknown length

15980 18644: contig of 2765 bp in length
18645 18744: gap of unknown length
18745 22176: contig of 3432 bp in length
22177 22276: gap of unknown length
22277 24448: contig of 2172 bp in length
24449 24548: gap of unknown length
24549 27217: contig of 2669 bp in length
27218 27317: gap of unknown length
27318 29554: contig of 2237 bp in length
29555 29654: gap of unknown length
29655 32154: contig of 2500 bp in length
32155 32254: gap of unknown length
32255 34292: contig of 2038 bp in length
34293 34392: gap of unknown length
34393 36336: contig of 1944 bp in length
36337 36436: gap of unknown length
36437 38645: contig of 2209 bp in length
38646 38745: gap of unknown length
38746 40703: contig of 1958 bp in length
40704 40803: gap of unknown length
40804 43062: contig of 2259 bp in length
43063 43162: gap of unknown length
43163 45360: contig of 2198 bp in length
45361 45460: gap of unknown length
45461 46677: contig of 1217 bp in length
46678 46777: gap of unknown length
46778 48470: contig of 1693 bp in length
48471 50423: contig of 1853 bp in length
50424 50523: gap of unknown length
50524 52685: contig of 2162 bp in length
52686 52785: gap of unknown length
52786 55045: contig of 2260 bp in length
55046 55145: gap of unknown length
55146 57896: contig of 2751 bp in length
57897 57996: gap of unknown length
57997 59874: contig of 1878 bp in length
59875 61349: contig of 1375 bp in length
61350 61449: gap of unknown length
61450 62557: contig of 1108 bp in length
62558 63943: contig of 1286 bp in length
63944 64043: gap of unknown length
64044 65504: contig of 1461 bp in length
65505 65604: gap of unknown length
65605 67142: contig of 1538 bp in length
67143 67242: gap of unknown length
67243 68565: contig of 1323 bp in length
68566 68665: gap of unknown length
68666 70653: contig of 1988 bp in length
70654 70753: gap of unknown length
70754 72145: contig of 1392 bp in length
72146 72245: gap of unknown length
72246 73583: contig of 1338 bp in length
73584 73683: gap of unknown length
73684 75163: contig of 1480 bp in length
75164 75263: gap of unknown length
75264 76816: contig of 1553 bp in length
76817 78916: gap of unknown length
78917 78994: contig of 1678 bp in length
78995 78694: gap of unknown length
78695 80054: contig of 1360 bp in length
80055 80154: gap of unknown length
80155 81845: contig of 1691 bp in length
81846 81945: gap of unknown length
81946 83461: contig of 1516 bp in length
83462 83561: gap of unknown length
83562 84551: contig of 1890 bp in length
84552 85551: gap of unknown length
85552 86618: contig of 1067 bp in length
86619 86718: gap of unknown length
86719 87882: contig of 1164 bp in length

* 87883 87982: gap of unknown length
 * 87983 89541: contig of 1559 bp in length
 * 89542 89641: gap of unknown length
 * 91013 91013: contig of 1372 bp in length
 * 91113 91113: gap of unknown length
 * 92643 92643: contig of 1530 bp in length
 * 92644 92743: gap of unknown length
 * 92744 93863: contig of 1120 bp in length
 * 93864 93863: gap of unknown length
 * 93964 94969: contig of 1006 bp in length
 * 95069 95069: gap of unknown length
 * 95070 96254: contig of 1185 bp in length
 * 96255 96355: gap of unknown length
 * 96355 97530: contig of 1176 bp in length
 * 97531 97630: gap of unknown length
 * 97631 98551: contig of 1020 bp in length
 * 98551 98750: gap of unknown length
 * 98751 99973: contig of 1223 bp in length
 * 99974 100074 101735: contig of 1662 bp in length
 * 101736 101835: gap of unknown length
 * 101836 103153: contig of 1318 bp in length
 * 103154 103253: gap of unknown length
 * 103254 104352: contig of 1099 bp in length
 * 104353 104452: gap of unknown length
 * 104453 105861: contig of 1409 bp in length
 * 105862 107264: contig of 1303 bp in length
 * 107265 107364: gap of unknown length

Query Match 2.5%; Score 19; DB 2; Length 125018;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 CCAGACACACCGCTGTT 505
 Db 79957 CCAGACACACCGCTGTT 79939
 |||||

RESULT 60
 LOCUS AC008560 129641 bp DNA linear PRI 23-AUG-2001
 DEFINITION Homo sapiens chromosome 5 clone CTC-534A2, complete sequence.
 ACCESSION AC008560
 VERSION AC008560.6 GI:15281180
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 129641)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 129641)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 129641)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 129641)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (23-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Aug 23, 2001 this sequence version replaced gi:10567844.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center

www-shgc.stanford.edu
 Quality: Phrap Quality >40 99.5% of Sequence;
 Estimated Total Number of Errors is 0.7.
 STS Content:
 WI-15163 G21073
 SHGC-33865 G29536.

FEATURES
 source
 1..129641
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTC-534A2"
 BASE COUNT 40288 a 23498 c 24388 g 41467 t
 ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 129641;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 AGTGCTTCAGAAATCCAGT 208
 Db 118144 AGTGCTTCAGAAATCCAGT 118162
 |||||

RESULT 61
 LOCUS AC016620 129856 bp DNA linear PRI 03-OCT-2001
 DEFINITION Homo sapiens chromosome 5 clone CTD-2318A17, complete sequence.
 ACCESSION AC016620
 VERSION AC016620.7 GI:15887294
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 129856)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 129856)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 129856)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (28-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 129856)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Oct 3, 2001 this sequence version replaced gi:10334873.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www-shgc.stanford.edu
 Quality: Phrap Quality >40 99.8% of Sequence;
 Estimated Total Number of Errors is 0.2.
 STS Content:
 SHGC-33865 G29536
 SHGC-140797 G63024.

FEATURES
 source
 1..129856
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTD-2318A17"
 BASE COUNT 39491 a 23663 c 25197 g 41505 t
 ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 129856;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 AGTGCTTCAGAAATCCAGT 208
 |||||
 DB 22506 AGTGCTTCAGAAATCCAGT 22524

RESULT 62
 AC012031/C
 LOCUS
 DEFINITION Homo sapiens chromosome 3 clone RP11-91K9, *** SEQUENCING IN
 PROGRESS ***; 25 unordered pieces.

AC012031
 AC012031.11 GI:15723536
 HTG: HTGS_PHASE1.
 human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 138725)
 Alstrooks,S.L., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
 Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,K., Jolivet,S.,
 Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korhavi,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Louised,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mavingey,E., McLeod,M.P., Meador,M.,
 Mei,O., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabhat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogulu,M., Okwuonu,C., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
 Slisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 138725)

Worley,K.C.

Direct Submission

Submitted (19-OCT-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Sep 23, 2001 this sequence version replaced gi:8699977.

----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc.help@bcm.tmc.edu
 ----- Project Information
 Center project name: RP11-91K9
 Center clone name: RP11-91K9
 ----- Summary Statistics
 Sequencing vector: M13: L08821
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 114463 bases at least Q40
 Consensus quality: 128469 bases at least Q30
 Consensus quality: 136159 bases at least Q20
 Estimated insert size: 138198; sum-of-contigs estimation
 Estimated insert size: 139891; agarose-fp estimation
 Quality coverage: 2.8x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 16124: contig of 16124 bp in length
 * 16224: gap of unknown length
 * 25186: contig of 8962 bp in length
 * 25187: gap of unknown length
 * 25287: gap of unknown length
 * 34462: contig of 9176 bp in length
 * 34562: gap of unknown length
 * 43750: contig of 9188 bp in length
 * 43751: gap of unknown length
 * 43850: gap of unknown length
 * 43851: contig of 7416 bp in length
 * 51266: gap of unknown length
 * 51267: gap of unknown length
 * 58270: contig of 6904 bp in length
 * 58271: gap of unknown length
 * 58370: contig of 6459 bp in length
 * 64839: gap of unknown length
 * 64939: gap of unknown length
 * 72181: contig of 7242 bp in length
 * 72182: gap of unknown length
 * 72281: gap of unknown length
 * 78119: contig of 5838 bp in length
 * 78220: gap of unknown length
 * 78220: gap of unknown length
 * 81638: contig of 3419 bp in length
 * 81639: gap of unknown length
 * 81739: contig of 4125 bp in length
 * 85863: gap of unknown length
 * 85864: gap of unknown length
 * 91272: contig of 5309 bp in length
 * 91273: gap of unknown length
 * 91372: gap of unknown length
 * 96272: gap of unknown length
 * 96273: gap of unknown length
 * 101124: contig of 4752 bp in length
 * 101125: gap of unknown length
 * 101224: gap of unknown length
 * 106567: contig of 5343 bp in length
 * 106568: gap of unknown length
 * 106568: contig of 4354 bp in length
 * 11021: contig of 4354 bp in length
 * 111022: gap of unknown length
 * 111121: gap of unknown length
 * 111122: contig of 3064 bp in length
 * 114185: gap of unknown length
 * 114186: gap of unknown length
 * 114866: gap of unknown length
 * 117972: contig of 3687 bp in length
 * 117973: gap of unknown length
 * 121659: contig of 3587 bp in length
 * 121660: gap of unknown length
 * 121759: gap of unknown length
 * 124487: contig of 2728 bp in length
 * 124488: gap of unknown length
 * 124587: gap of unknown length
 * 124588: gap of unknown length
 * 127179: gap of unknown length
 * 130129: contig of 2851 bp in length
 * 130229: gap of unknown length

* 130230 133622: contig of 3393 bp in length
 * 133623 133722: gap of unknown length
 * 133723 136582: contig of 2860 bp in length
 * 136583 136582: gap of unknown length
 * 136683 138725: contig of 2043 bp in length.

FEATURES

source

1. 138725
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /clone="RP11-91K9"
 /clone="RP11-91K9"

BASE COUNT 43006 a 26285 c 26855 g 40143 t 2436 others
 ORIGIN

Query Match 2.5%: Score 19; DB 2; Length 138725;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 AAGCTGAGAAATTCCTGAA 258

|||||

Db 90882 AAGCTGAGAAATTCCTGAA 90864

RESULT 63

AP003809

LOCUS

DEFINITION Oryza sativa chromosome 7 clone OJ116.D12, linear HTG 04-JUL-2001
 PROGRESS ***, in ordered pieces.

ACCESSION

AP003809

VERSION

AP003809.1 GI:14595149

KEYWORDS

HTG; HTGS_PHASE2.

SOURCE

Oryza sativa

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

Sasaki, T., Matsumoto, T. and Yamamoto, K.

AUTHORS

TITLE

Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
 clone:OJ116.D12

JOURNAL

Published Only in Database (2001) In press

2 (bases 1 to 139629)

Sasaki, T., Matsumoto, T. and Yamamoto, K.

Direct Submission

Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program; Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@abrc.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/.

Tel:81-298-38-7441, Fax:81-298-38-7468)

The nucleotide sequence of this BAC clone was generated by
 combining Monsanto and RGP-Japan sequencing data.
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source

1. 139629
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="7"
 /clone="OJ116.D12"

BASE COUNT 40427 a 29835 c 29223 g 39787 t 357 others
 ORIGIN

Query Match

2.5%: Score 19; DB 2; Length 139629;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 GCAACTGCCAAGCAGAAG 232

|||||

Db 48405 GCAACTGCCAAGCAGAAG 48424

RESULT 64

AP004335

LOCUS

DEFINITION Oryza sativa chromosome 7 clone P0483G08, linear HTG 15-NOV-2001
 PROGRESS ***, in ordered pieces.

ACCESSION

AP004335

VERSION

AP004335.1 GI:16930120

KEYWORDS

HTG; HTGS_PHASE2.

SOURCE

Oryza sativa

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

Sasaki, T., Matsumoto, T. and Yamamoto, K.

AUTHORS

TITLE

Direct Submission

Submitted (14-NOV-2001) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program; Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@abrc.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/.

Tel:81-298-38-7441, Fax:81-298-38-7468)

NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source

1. 139848
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="7"
 /clone="P0483G08"

BASE COUNT 41254 a 29682 c 29071 g 39841 t

ORIGIN

Query Match

2.5%: Score 19; DB 2; Length 139848;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 GCAACTGCCAAGCAGAAG 232

|||||

Db 104339 GCAACTGCCAAGCAGAAG 104357

RESULT 65

AC011130/C

LOCUS

DEFINITION Homo sapiens chromosome 18 clone RP11-268I3 map 18, WORKING DRAFT
 PROGRESS ***, in ordered pieces.

ACCESSION

AC011130

VERSION

AC011130.4 GI:8072502

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 142388)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-26813
2 (bases 1 to 142388)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Batwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardy,J., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Leibach,J., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7133525.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2366
Center clone name: 268.1.3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 133379 bases at least Q40
Consensus quality: 137445 bases at least Q30
Consensus quality: 139434 bases at least Q20
Insert size: 137000; agarose-fp
Insert size: 141088; sum-of-ctngs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctngs

TITLE
JOURNAL
COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1073: contig of 1073 bp in length
* 1074 1173: gap of 100 bp
* 1174 2606: contig of 1433 bp in length
* 2607 2706: gap of 100 bp
* 2707 4915: contig of 2209 bp in length
* 4916 5015: gap of 100 bp
* 5016 6466: contig of 1453 bp in length
* 6469 6568: gap of 100 bp
* 6569 10588: contig of 4020 bp in length
* 10589 10688: gap of 100 bp
* 10689 15082: contig of 4394 bp in length
* 15083 15182: gap of 100 bp
* 15183 20848: contig of 5666 bp in length
* 20849 20948: gap of 100 bp
* 20949 27902: contig of 6954 bp in length
* 27903 28002: gap of 100 bp
* 28003 36222: contig of 8220 bp in length
* 36223 36322: gap of 100 bp
* 36323 45886: contig of 9564 bp in length
* 45887 45986: gap of 100 bp

* 45987 57780: contig of 11794 bp in length
* 57781 57880: gap of 100 bp
* 57881 70671: contig of 12791 bp in length
* 70672 70771: gap of 100 bp
* 70772 95778: contig of 25007 bp in length
* 95779 95878: gap of 100 bp
* 95879 142388: contig of 46510 bp in length.
FEATURES
Location/Qualifiers
Source
1. 142388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-26813"
/clone.lib="RP11-11 Human Male BAC"
1. 1073
/note="assembly_fragment"
1174..2606
/note="assembly_fragment"
2707..4915
/note="assembly_fragment"
5016..6468
/note="assembly_fragment"
clone_end:SP6
vector_side:right
6569..10588
/note="assembly_fragment"
10689..15082
/note="assembly_fragment"
15183..20848
/note="assembly_fragment"
20949..27902
/note="assembly_fragment"
28003..36222
/note="assembly_fragment"
36323..45886
/note="assembly_fragment"
45987..57780
/note="assembly_fragment"
57881..70671
/note="assembly_fragment"
clone_end:T7
vector_side:right
70772..95778
/note="assembly_fragment"
95879..142388
/note="assembly_fragment"
BASE COUNT 44920 a 27060 c 26345 g 42758 t 1305 others
ORIGIN

Query Match 2.5% Score 19; DB 2; Length 142388;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 AATCTGGCAATGACCTAA 89
|||||
DB 105952 AATCTGGCAATGACCTAA 105934

RESULT 66
AC103905

LOCUS
DEFINITION
Canis familiaris clone RP81-234H11, WORKING DRAFT SEQUENCE, 4
linear HTG 30-NOV-2001
unordered pieces.

ACCESSION AC103905

VERSION AC103905.1 GI:17155058

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE dog.

ORGANISM Canis familiaris

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 142603)

AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Ho, S.-L., Ito, J.R., Karlins, E., Leric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B., Masello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y., Stantripoop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgueon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Zhang, L.H., and Green, E.D.

TITLE NISC Comparative Sequencing Initiative

REFERENCE Unpublished

AUTHORS 2 (bases 1 to 142603)

JOURNAL Green, E.D.

TITLE Direct Submission

JOURNAL Submitted (30-NOV-2001) NIH Intramural Sequencing Center, 8717 Government Circle, Gaithersburg, MD 20877, USA

COMMENT ----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_mouse@hgrl.nih.gov
----- Project Information
Center project name: cpv
Center clone name: 234H11
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 141458 bases at least Q40
Consensus quality: 141778 bases at least Q30
Consensus quality: 141869 bases at least Q20
Insert size: 91000; agarose-fp
Insert size: 142303; sum-of-contigs
Quality coverage: 13.58x in Q20 bases; agarose-fp
Quality coverage: 8.69x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 8113: contig of 8113 bp in length
8114 8213: gap of unknown length
8214 35526: contig of 27313 bp in length
35527 35626: gap of unknown length
35627 78480: contig of 42854 bp in length
78481 78580: gap of unknown length
78581 142603: contig of 64023 bp in length.

FEATURES
source
1. .142603
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone="RP81-234H11"
/clone_lib="RP81"
misc_feature
1. .8113
/note="assembly_fragment"
8214. .35526
/note="assembly_fragment"
35627. .78480
/note="assembly_fragment"
clone_end:T7
vector_side:left
78581. .142603
/note="assembly_fragment"
clone_end:SP6
vector_side:right
BASE COUNT 42402 a 28609 c 28267 g 43025 t 300 others
ORIGIN

Query Match 2.5% Score 19: DB 2: Length 142603;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 628 CTCCTGCTGTCAGAGGAG 646
|||||
Db 105204 CTCCTGCTGTCAGAGGAG 105222

RESULT 67
AC016611/C
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2197B7, WORKING DRAFT SEQUENCE,
11 ordered pieces.
AC016611
VERSION 1.6 GI-9256350
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
2 (bases 1 to 145899)
Direct Submission
Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7710184.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 703228
Center clone name: CITB-HL_2197B7

Summary Statistics
Consensus quality: 134277 bases at least Q40
Consensus quality: 142520 bases at least Q30
Consensus quality: 143768 bases at least Q20
Estimated insert size: 145000; pulse field gel estimation
Quality coverage: 5.12 in Q20 bases; sum-of-contigs estimation
Quality coverage: 5.1 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. Gaps between the pieces are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 730: contig of 730 bp in length
731 830: gap of unknown length
831 2648: contig of 1818 bp in length
2649 2748: gap of unknown length
2749 30649: contig of 27901 bp in length
30650 30749: gap of unknown length
30750 43422: contig of 12673 bp in length
43423 43523: gap of unknown length
43524 55335: contig of 11813 bp in length
55336 55436: gap of unknown length
55437 86072: contig of 30636 bp in length
86073 86171: gap of unknown length
86172 115132: contig of 28960 bp in length
115133 115231: gap of unknown length
115232 122471: contig of 7239 bp in length
122472 122570: gap of unknown length

```

* 122571 124854: contig of 2284 bp in length
* 124855 124954: gap of unknown length
* 124955 131924: contig of 6970 bp in length
* 131925 132024: gap of unknown length
* 132025 145899: contig of 13875 bp in length.
      Location/Qualifiers
        1. 145899
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /chromosome="5"
          /clone="CTD-2197B7"
          /clone_lib="Caltech human BAC library D"
BASE COUNT 46354 a 27496 c 26688 g 44356 t 1005 others
ORIGIN

Query Match      2.5%  Score 19;  DB 2;  Length 145899;
Best Local Similarity 100.0%;  Pred. No. 34;
Matches 19;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 190 AGTGGTCAGAAATCCACT 208
      |||||
Db 2884 AGTGGTCAGAAATCCACT 2866

RESULT 68
AC106459
LOCUS
DEFINITION
AC106459
AC106459.1 GI:18138981
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Muzny D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,
Alsbrooks S.L., Amarantunga H.C., Are J.R., Banks T., Barbara J.,
Benton J., Binage K., Blankenburg K., Bonnin D., Bouck J.,
Bowle S., Brieve M., Brown E., Brown M., Bryant N.P., Buhay C.,
Burke P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F.,
Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,
Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,
Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,
Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C.,
Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,
Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,
Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,
Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,
Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C.,
Hollins B., Homs F., Howard S., Huber J., Hulyk S., Hume J.,
Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S.,
Joudan S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,
Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,
Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W.,
Loulsegh H., Lozardo R.J., Lu X., Lucier A., Lucier R., Luna R.,
Ma J., Maheshwari M., Mapa P., Martin R., Martindale A.,
Martinez E., Massey E., Mawhney E., McLeod M.P., Meador M.,
Mei G., Metzger M., Miner G., Miner Z., Mitchell T., Mohabbat K.,
Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N.,
Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokenwo S.,
Ogih M., Okwuonu G., Oragunye N., Oviedo R., Pace A., Payton B.,
Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,
Quiles M., Ren Y., Rives M., Rojas A., Rojibokan I., Rolfe M.,
Ruiz S., Savary G., Scher S., Scott G., Shen H., Shoostari N.,
Sison I., Sodergren E., Somaite A., Tabor P., Sparks A., Stanley H.,
Stone H., Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K.,
Tang H., Tansey J., Taylor C., Taylor T., Telford B., Thomas N.,

```

Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 151548)
Worley, K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GKJ
Center clone name: CH230-207G10

Summary Statistics
Assembly program: Phrap; version 0.990329First call to findPrapList
Consensus quality: 130235 bases at least Q40
Consensus quality: 138462 bases at least Q30
Consensus quality: 143966 bases at least Q20
Estimated insert size: 134069; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently consists of 63 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 6827: contig of 6827 bp in length
6828 6927: gap of unknown length
6928 10296: contig of 3369 bp in length
10297 10396: gap of unknown length
10397 15480: contig of 5084 bp in length
15481 15880: gap of unknown length
15881 20702: contig of 5122 bp in length
20703 20802: gap of unknown length
20803 25073: contig of 4271 bp in length
25074 25173: gap of unknown length
25174 29656: contig of 4483 bp in length
29657 29756: gap of unknown length
29757 34158: contig of 4402 bp in length
34159 34258: gap of unknown length
34259 39340: contig of 5082 bp in length
39341 39440: gap of unknown length
39441 43252: contig of 3812 bp in length
43253 43353: gap of unknown length
43353 46447: contig of 3095 bp in length
46448 46547: gap of unknown length
46548 49762: contig of 3215 bp in length
49763 49863: gap of unknown length
49863 52060: contig of 2198 bp in length
52061 52160: gap of unknown length
52161 55316: contig of 3156 bp in length
55317 55417: gap of unknown length
55417 57690: contig of 2274 bp in length
57691 57790: gap of unknown length
57791 60306: contig of 2516 bp in length
60307 60407: gap of unknown length
60407 62706: contig of 2300 bp in length
62707 62806: gap of unknown length

62807 * 65749: contig of 2943 bp in length
 65750 * 65849: gap of unknown length
 65850 * 69387: contig of 3538 bp in length
 69388 * 69487: gap of unknown length
 69488 * 71954: contig of 2467 bp in length
 71955 * 72054: gap of unknown length
 72055 * 74901: contig of 2847 bp in length
 74902 * 75001: gap of unknown length
 75003 * 78717: contig of 3716 bp in length
 78718 * 78817: gap of unknown length
 78818 * 81150: contig of 2333 bp in length
 81151 * 81250: gap of unknown length
 81251 * 83005: contig of 1755 bp in length
 83006 * 83105: gap of unknown length
 83106 * 85134: contig of 2029 bp in length
 85135 * 85234: gap of unknown length
 85235 * 85754: contig of 2340 bp in length
 85755 * 87674: gap of unknown length
 87675 * 89777: contig of 2103 bp in length
 89778 * 89877: gap of unknown length
 89879 * 91833: contig of 1956 bp in length
 91834 * 91933: gap of unknown length
 91934 * 93738: contig of 1805 bp in length
 93739 * 93838: gap of unknown length
 93839 * 96413: contig of 2575 bp in length
 96414 * 96514: gap of unknown length
 96515 * 99247: contig of 2734 bp in length
 99248 * 99347: gap of unknown length
 99349 * 101218: contig of 1871 bp in length
 101219 * 101318: gap of unknown length
 101319 * 102836: contig of 1518 bp in length
 102837 * 102937: gap of unknown length
 102938 * 104287: contig of 1351 bp in length
 104288 * 104387: gap of unknown length
 104389 * 106347: contig of 1960 bp in length
 106348 * 106447: gap of unknown length
 106448 * 107589: contig of 1142 bp in length
 107590 * 107689: gap of unknown length
 107690 * 108939: contig of 1250 bp in length
 108940 * 109039: gap of unknown length
 109040 * 110174: contig of 1135 bp in length
 110175 * 110274: gap of unknown length
 110275 * 112147: contig of 1873 bp in length
 112148 * 112447: gap of unknown length
 112448 * 114101: contig of 1854 bp in length
 114102 * 114201: gap of unknown length
 114202 * 115932: contig of 1731 bp in length
 115933 * 116032: gap of unknown length
 116033 * 117806: contig of 1674 bp in length
 117807 * 117903: gap of unknown length
 117904 * 119003: contig of 2097 bp in length
 119004 * 120003: gap of unknown length
 120004 * 121033: contig of 1030 bp in length
 121034 * 121134: gap of unknown length
 121135 * 122822: contig of 1689 bp in length
 122823 * 122922: gap of unknown length
 122923 * 123957: contig of 1035 bp in length
 123958 * 124057: gap of unknown length
 124058 * 125138: contig of 1081 bp in length
 125139 * 125238: gap of unknown length
 125239 * 127008: contig of 1770 bp in length
 127009 * 127108: gap of unknown length
 127109 * 128481: contig of 1373 bp in length
 128482 * 128581: gap of unknown length
 128583 * 130196: contig of 1615 bp in length
 130197 * 130296: gap of unknown length
 130297 * 131477: contig of 1181 bp in length
 131478 * 131577: gap of unknown length
 131578 * 133181: contig of 1604 bp in length
 133182 * 133261: gap of unknown length
 133262 * 134538: contig of 1257 bp in length
 134539 * 134638: gap of unknown length
 134639 * 136503: contig of 1865 bp in length

Query Match 2.5% Score 19; DB 2; Length 151548;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 ACATTTTGGGTTCAAAAT 312

|||||TTTTTTTTTTTTTT

Db 132934 ACATTTTGGGTTCAAAAT 132952

RESULT 69

AC019246

LOCUS Homo sapiens, clone RP11-2K18, complete sequence.

DEFINITION AC019246 154984 bp DNA linear PRI 01-MAY-2001

AC019246

AC019246.6 GI:13899425

HTG

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 154984)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens, clone RP11-2K18

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 154984)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F., Boguslavsky, L., Bouckghalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lied, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 154984)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckghalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S.,

Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zemlek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 1, 2001 this sequence version replaced gi:11415189.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1547
 Center clone name: 2_K_18

FEATURES

source	Location/Qualifiers	repeat_region
1..154984	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="Rpl1-2X18"	
	/clone_lib="RPL1-11 Human Male BAC"	
	complement(2619..2763)	
	/rpt_family="MSTD"	
	2764..2786	
	/rpt_family="(CA)n"	
	complement(2787..2982)	
	/rpt_family="MSTD"	
	complement(2983..4612)	
	/rpt_family="MSTD-int"	
	complement(4615..4848)	
	/rpt_family="MSTD"	
	complement(4858..4940)	
	/rpt_family="AluJ/FRAM"	
	6568..6590	
	/rpt_family="(T)n"	
	complement(8741..9003)	
	/rpt_family="MIR"	
	10197..10507	
	/rpt_family="AluSg"	
	complement(10727..10798)	
	/rpt_family="L2"	
	11008..11313	
	/rpt_family="AluJb"	
	11815..11843	
	/rpt_family="(TG)n"	
	11966..12003	
	/rpt_family="AT-rich"	
	12445..12465	
	/rpt_family="AT-rich"	
	13111..13337	
	/rpt_family="L1MA9"	
	13625..13936	
	/rpt_family="AluSx"	
	14090..14152	
	/rpt_family="AT-rich"	
	14581..14955	
	/rpt_family="MLT2C1"	
	14989..15019	
	/rpt_family="AT-rich"	
	15955..18040	
	/rpt_family="L1MD3"	
	18039..19176	
	/rpt_family="L1MD3"	
	19325..19442	
	/rpt_family="LTR16C"	
	complement(20192..20250)	
	/rpt_family="MERSB"	
	20627..20714	
	/rpt_family="AT-rich"	
	complement(21552..22050)	
	/rpt_family="L1MA6"	
	22232..22563	
	/rpt_family="MLT2C1"	
	22564..22598	
	/rpt_family="(TC)n"	
	22598..22628	
	/rpt_family="(TG)n"	
	22599..22630	
	/rpt_family="MLT2C1"	
	22631..23392	
	/rpt_family="Tigger2"	
	23393..23422	
	/rpt_family="(TAAA)n"	
	23423..23528	
	/rpt_family="Tigger2"	
	23529..23838	
	/rpt_family="AluSc"	
	23839..25637	
	/rpt_family="Tigger2"	
	25644..25918	
	/rpt_family="AluSx"	
	25922..25950	
	/rpt_family="(CA)n"	
	25951..26014	
	/rpt_family="MLT2C1"	
	26594..26636	
	/rpt_family="AT-rich"	
	26644..26848	
	/rpt_family="Charliel"	
	26860..27160	
	/rpt_family="AluJo"	
	27232..27511	
	/rpt_family="Charliel"	
	27513..27648	
	/rpt_family="Charliel"	
	28787..29171	
	/rpt_family="L2"	
	complement(29864..30029)	
	/rpt_family="THE1B"	
	30236..30355	
	/rpt_family="L2"	
	30707..31010	
	/rpt_family="AluSg1"	
	complement(31072..31193)	
	/rpt_family="MERSB"	
	31220..31243	
	/rpt_family="AT-rich"	
	31350..31391	
	/rpt_family="AT-rich"	
	complement(31481..31882)	
	/rpt_family="MLT1A1"	
	complement(32650..32994)	
	/rpt_family="THE1C"	
	33040..33339	
	/rpt_family="AluY"	
	complement(34429..34689)	
	/rpt_family="MERS4A"	
	35768..37107	
	/rpt_family="L1PA6"	
	37463..37749	
	/rpt_family="AluJo"	
	complement(38887..39250)	
	/rpt_family="MLT1C"	
	39989..40010	
	/rpt_family="AT-rich"	
	40189..40216	
	/rpt_family="AT-rich"	
	40351..40663	
	/rpt_family="AluYb8"	
	41000..41062	
	/rpt_family="(TTTA)n"	
	41070..41655	


```

RESULT 71
AP003101
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-30J7, complete
sequences.
ACCESSION AP003101
VERSION AP003101.2 GI:17939959
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-30J7.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 157959)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Dec 19, 2001 this sequence version replaced gi:1275508.
FEATURES
Source
location/Qualifiers
1..157959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-30J7"
BASE COUNT 40406 a 38903 c 38805 g 39845 t
ORIGIN
Query Match 2.5%; Score 19; DB 9; Length 157959;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 230 AGGCCCGAGGAGCTGAGA 248
|||||
DB 50592 AGGCCCGAGGAGCTGAGA 50610

RESULT 72
AC022644
LOCUS
DEFINITION Homo sapiens clone RP11-28A20, WORKING DRAFT SEQUENCE, 32 unordered
pieces.
ACCESSION AC022644
VERSION AC022644.2 GI:7139692
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 160355)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-28A20
JOURNAL Unpublished
REFERENCE 2. (bases 1 to 160355)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,K., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Garayna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,C., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Menues,L., Mortow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Kiley,R., Rothman,D.,

```

Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 1, 2000 this sequence version replaced gi:6910664.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1998-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L4736
 Center clone name: 28_A20
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 141286 bases at least Q40
 Consensus quality: 151099 bases at least Q30
 Consensus quality: 154621 bases at least Q20
 Insert size: 176000; agarose-fp
 Insert coverage: 157255; sum-of-contigs
 Quality coverage: 3.1 in Q20 bases; agarose-fp
 Quality coverage: 3.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1064: contig of 1064 bp in length
 1065 1164: gap of 100 bp
 1165 2682: contig of 1518 bp in length
 2683 2782: gap of 100 bp
 2783 3977: contig of 1195 bp in length
 3978 4077: gap of 100 bp
 4078 5829: contig of 1752 bp in length
 5830 5929: gap of 100 bp
 5930 7859: contig of 1930 bp in length
 7860 7959: gap of 100 bp
 7960 9560: contig of 1601 bp in length
 9561 9660: gap of 100 bp
 9661 11027: contig of 1367 bp in length
 11028 11127: gap of 100 bp
 11128 13507: contig of 2380 bp in length
 13508 13607: gap of 100 bp
 13608 15417: contig of 1810 bp in length
 15418 15517: gap of 100 bp
 15518 18337: contig of 2820 bp in length
 18338 18437: gap of 100 bp
 18438 20702: contig of 2265 bp in length
 20703 20802: gap of 100 bp
 20803 23501: contig of 2699 bp in length
 23502 23601: gap of 100 bp
 23602 26976: contig of 3375 bp in length
 26977 27076: gap of 100 bp
 27077 29993: contig of 2917 bp in length
 29994 30093: gap of 100 bp
 30094 33031: contig of 2938 bp in length
 33032 33131: gap of 100 bp
 33132 35424: contig of 2293 bp in length
 35425 35524: gap of 100 bp
 35525 38714: contig of 3190 bp in length
 38715 38814: gap of 100 bp

Quality coverage: 10.27x in Q20 bases; sum-of-contigs Quality coverage: 9.26x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 3614: contig of 3614 bp in length
 3615 3714: gap of 100 bp
 3715 21865: contig of 18151 bp in length
 21866 21965: gap of 100 bp
 21966 26393: contig of 4428 bp in length
 26394 26493: gap of 100 bp
 26494 66846: contig of 40353 bp in length
 66847 66946: gap of 100 bp
 66947 94641: contig of 27695 bp in length
 94642 94741: gap of 100 bp
 94742 131462: contig of 36721 bp in length
 131463 131562: gap of 100 bp
 131563 160725: contig of 29163 bp in length.

FEATURES

source

Location/Qualifiers
 1..160725
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="q12.3-13.32"
 /clone="RP5-919B11"
 /clone_lib="RPC1-5"
 misc_feature
 1..3614
 /note="assembly_fragment:00088"
 clone_end:T7
 vector_side:left
 3715..21865
 /note="assembly_fragment:00888"
 21966..26393
 /note="assembly_fragment:02354"
 26494..66846
 /note="assembly_fragment:02994"
 66947..94641
 /note="assembly_fragment:03237"
 94742..131462
 /note="assembly_fragment:03896"
 131563..160725
 /note="assembly_fragment:05273"
 37797 a 42852 c 42211 g 37262 t 603 others

BASE COUNT 37797 a 42852 c 42211 g 37262 t 603 others
 ORIGIN

Query Match 2.58; Score 19; DB 2; Length 160725;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 AAGGCCAGGAGCTGACA 248
 |||||
 Db 48305 AAGGCCAGGAGCTGACA 48323

RESULT 74

AC092812/C AC092812 160903 bp DNA linear PRI 09-NOV-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-536F12, complete sequence.
 DEFINITION AC092812 AL360299
 ACCESSION AC092812.2 GI:16874859
 VERSION HTG.
 KEYWORDS SOURCE
 ORGANISM human.
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 160903)

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and Haugen,E.D.
 Direct Submission
 2 (bases 1 to 160903)
 Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and Haugen,E.D.
 Direct Submission
 Submitted (28-JUL-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 160903)
 Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and Haugen,E.D.
 Direct Submission
 Submitted (09-NOV-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 On Nov 9, 2001 this sequence version replaced gi:15027765.

Center: University of Washington Genome Center

Center Code: UWGC

Web site: <http://www.genome.washington.edu>

Contact: uwgctgs@u.washington.edu

Drafting Center: SC

----- Project Information

Center project name: chr-1

Center clone name: RP11-536F12 (sc0172)

----- Summary Statistics

Sequencing vector: Plasmid; 46% of reads

Chemistry: Dye-terminator ET; 89% of reads

Chemistry: Dye-terminator Big Dye; 11% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 160823 bases at least Q40

Consensus quality: 160891 bases at least Q30

Consensus quality: 160903 bases at least Q20

Insert size: 151225; 11.5% error; agarose-fp

Insert size: 160907; sum-of-contigs

Quality coverage: 9.2x in Q20 bases; agarose-fp

Quality coverage: 8.7x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:

5': Mapping in progress

3': RP4-798D13 (UWGC:sc0214) AL354679

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

GenBank flat file format but are available as part

of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:

all regions were either double-stranded or sequenced with an

alternate chemistry or covered by high quality data (i.e., Phred

quality >= 30); an attempt was made to resolve all sequencing

problems, such as compressions and repeats; all regions were

covered by at least one plasmid subclone or more than one M13

subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest

fingerprinting. Comparison of the experimentally derived digest

fragments with sequence-predicted fragments is given below.

The electronically-digested sequence consists of both insert and

vector, in order to accurately represent the entire circular BAC.

Small fragments below a variable cutoff (approximately 400-600 bp)

are not resolved in the fingerprint and hence do not appear

in the table. There are no significant remaining discrepancies

between the experimental and predicted values. Uniquely ordered

fragments are separated by dashed lines.

EcoRI		HindIII		BglII	
SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt
8696	8836	3639	3399	6080	5938
6	<800	6382	6500	2067	2062
112	<800	512	<800	5523	5938
597	<800	449	<800	3799	3750
3420	3383	13872	13818	1601	1570
1878	1978	706	<800	1030	1075
880	864	580	<800	10441	10399
1945	1978	4008	3980	11548	11464
2022	1978	5403	5371	3389	3315
340	<800	121	<800	9805	9848
4775	4742	7032	7262	7494	7550
1329	1356	411	<800	414	<800
6514	6537	4854	4777	1071	1075
36	<800	2905	2988	105	<800
950	946	6508	6500	9	<800
4222	4219	4264	4190	568	<800
11608	11557	5707	5684	9583	9472
2238	2255	9192	9117	321	<800
2857	2886	2796	2792	810	817
541	<800	3450	3399	2273	2316
7413	7453	6316	6500	304	<800
1407	1356	3465	3399	1721	1688
2277	2255	2987	2988	2872	2857
680	<800	1153	1159	416	<800
748	777	491	<800	1081	1075
5526	5430	3478	3663	863	879
1579	1543	2041	2089	3642	3620
12450	12275	58	<800	8901	8850
1733	1813	2424	2422	3982	3958
6906	6948	1843	1845	2980	2965
3743	3703	7363	7262	5978	5938
4555	4506	356	<800	560	<800
272	<800	212	<800	7263	7550
29	<800	575	<800	4849	4813

Query Match 2.5%; Score 19; DB 9; Length 160903;
 Best local Similarity 100.0%; Pred. NO. 35;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 CATGATTGGCCAGATTATC 433
 |||||
 DB 96922 CATGATTGGCCAGATTATC 96904

RESULT 75

AL157702

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human DNA sequence from clone Rp11-18816 on chromosome 9 Contains a novel pseudogene, complete sequence.

AL157702.10 GI:10086046

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 160990)

Rand.V.

Direct Submission

Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Sep 11, 2000 this sequence version replaced gi:10039651.

During sequence assembly data is compared from overlapping clones, where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>

This sequence is the entire insert of clone RP11-18B16 The true left end of clone RP11-53478 is at 158579 in this sequence. The true right end of clone RP11-168K11 is at 73774 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-18B16 is from the library RPc1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

FEATURES

```

1. 160990
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-18B16"
/clone_lib="RPC1-11.1"
6. 679
/note="match: GSS: Em:AQ115227"
400. 500
/note="MER5A repeat: matches 70. 187 of consensus"
520. 997
/note="MER31A repeat: matches 1. 484 of consensus"
2078. 2336
/note="HA1ib repeat: matches 109. 397 of consensus"
complement(2205. 2683)
/note="match: GSS: Em:AQ209886"
complement(2220. 2693)
/note="match: GSS: Em:AQ674620"
2389. 2624
/note="MIR repeat: matches 13. 252 of consensus"
2672. 2890
/note="HA1ib repeat: matches 595. 1751 of consensus"
2893. 3200
/note="AluSp repeat: matches 5. 311 of consensus"
3134. 3538
/note="match: STS: Em:G13893"
3786. 4078
/note="AluX repeat: matches 3. 295 of consensus"
4352. 4645
/note="L2 repeat: matches 2816. 3114 of consensus"
6123. 6180
/note="L2 repeat: matches 3202. 3257 of consensus"
6234. 6336
/note="MIR repeat: matches 18. 125 of consensus"
6786. 6842
/note="MR13 repeat: matches 115. 171 of consensus"
6928. 7268
/note="L2 repeat: matches 2808. 3191 of consensus"
7016. 7389
/note="match: GSS: Em:AQ19774"
7420. 7549
/note="FLAM_C repeat: matches 2. 132 of consensus"
8109. 8315

```

```

repeat_region 24383..24676
/note="AluSq repeat: matches 1..296 of consensus"
repeat_region 24677..24728
/note="MIR repeat: matches 9..61 of consensus"
repeat_region 25049..25143
/note="L2 repeat: matches 3213..3312 of consensus"
repeat_region 25126..25622
/note="L2 repeat: matches 2449..3000 of consensus"
repeat_region 25677..25968
/note="L2 repeat: matches 2961..3272 of consensus"
repeat_region 26013..26279
/note="AluJG repeat: matches 60..312 of consensus"
repeat_region 26504..26747
/note="L2 repeat: matches 2337..2579 of consensus"
repeat_region 26935..27308
/note="THELC repeat: matches 1..375 of consensus"
repeat_region 27693..27866
/note="L2 repeat: matches 2788..2973 of consensus"
repeat_region 28228..28705
/note="239 copies 2 mer 99 58% conserved"
repeat_region 29587..29750
/note="MIR repeat: matches 17..190 of consensus"
repeat_region 29786..29880
/note="MIR repeat: matches 46..145 of consensus"
misc_feature 29827..30213
/note="match: GSS: Em:AQ092683"
repeat_region 30177..30359
/note="MIR repeat: matches 9..208 of consensus"
repeat_region 31865..31943
/note="MIR repeat: matches 74..153 of consensus"
repeat_region 32096..32256
/note="MER3 repeat: matches 1..140 of consensus"
repeat_region 32257..32557
/note="AluSp repeat: matches 5..306 of consensus"
repeat_region 32558..32616
/note="MER3 repeat: matches 140..202 of consensus"
repeat_region 32750..32877

```

```

Query Match          2.5%; Score 19; DB 9; Length 160990;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 721 GCACAAACCCAGCTGCAA 739
    ||||||
Db 76321 GCACAAACCCAGCTGCAA 76339

```

```

Search completed: August 6, 2002, 20:07:27
Job time: 9651 sec

```

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.
 OM nucleic - nucleic search, using sw model
 Run on: August 6, 2002, 17:25:36 ; Search time 1700.58 Seconds
 (without alignments)
 5920.761 Million cell updates/sec
 Title: US-10-020-139-l_COPY_48_793
 Perfect score: 746
 Sequence: 1 GATGCTTCAGCTTGGAAAC.....AACCCAGCTGCAACCCCTCA 746

Scoring table: OLIGO_NUC
 Gapop 60.0 , Gapext 60.0
 Searched: 13736207 seqs, 6748477542 residues
 Word size: 0
 Total number of hits satisfying chosen parameters: 27472414
 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Listing first 250 summaries

Database: EST.*
 1: em_estba:*
 2: em_esthum:*
 3: em_estnu:*
 4: em_estnu:*
 5: em_estov:*
 6: em_estpl:*
 7: em_estro:*
 8: em_hic:*
 9: gb_esti:*
 10: gb_esti:*
 11: gb_hic:*
 12: gb_gss:*
 13: em_gss_hum:*
 14: em_gss_inv:*
 15: em_gss_pla:*
 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	124	16.6	384	9	AI654622 wb48f11.x
2	50	6.7	117	9	BE074387 QV3-BT057
3	50	6.7	117	9	BE074388 QV3-BT057
4	50	6.7	118	9	BE074383 QV3-BT057
5	49	6.6	130	9	BE074390 QV3-BT057
6	38	5.1	117	9	BE074380 QV3-BT057
7	36	4.8	119	9	BE074379 QV3-BT057
8	33	4.4	119	9	BE074382 QV3-BT057
9	23	3.1	406	10	BF651530 274330 MA
10	21	2.8	270	12	A2320113
11	21	2.8	425	12	A2977289
12	21	2.8	912	10	BE733517
13	20	2.7	358	9	BE028471
14	20	2.7	456	9	BE028127
15	20	2.7	475	10	BE751494
16	20	2.7	553	10	BE751485 203654 MA
17	20	2.7	574	10	B1540670 454023 MA

18	20	2.7	578	9	AV614200
19	20	2.7	661	10	BI288757
20	20	2.7	690	9	AW914186
21	20	2.7	695	10	BE273587
22	20	2.7	879	10	BI103654
23	19	2.5	415	10	BE954362
24	19	2.5	421	9	AA679456
25	19	2.5	424	9	AW758389
26	19	2.5	472	10	BF418859
27	19	2.5	477	9	A1244882
28	19	2.5	478	10	BE944264
29	19	2.5	481	10	BG790803
30	19	2.5	528	12	A0759251
31	19	2.5	528	10	BM386473
32	19	2.5	536	10	EG383314
33	19	2.5	540	10	BF279171
34	19	2.5	563	9	A1731923
35	19	2.5	569	9	AW141069
36	19	2.5	601	12	A0453959
37	19	2.5	610	12	A2446530
38	19	2.5	683	12	AZ858721
39	19	2.5	694	12	B66149
40	19	2.5	739	12	AZ978692
41	19	2.5	781	10	BF275075
42	19	2.5	782	10	BF275073
43	19	2.5	796	10	BI829148
44	19	2.5	845	10	BF275344
45	19	2.5	858	12	BM363397
46	19	2.5	1095	10	BI462868
47	19	2.5	2343	11	AK013777
48	18	2.4	152	9	AA980784
49	18	2.4	239	9	BB055152
50	18	2.4	269	9	AA062149
51	18	2.4	275	9	AL119042
52	18	2.4	276	9	AW312313
53	18	2.4	296	10	W36972
54	18	2.4	321	10	BF290469
55	18	2.4	323	9	AW662992
56	18	2.4	323	9	BB491192
57	18	2.4	332	9	AA751893
58	18	2.4	335	9	AA671815
59	18	2.4	335	9	BB24526
60	18	2.4	336	9	AA350438
61	18	2.4	341	9	AA285185
62	18	2.4	343	9	AI372485
63	18	2.4	351	12	AQ918301
64	18	2.4	359	10	W36801
65	18	2.4	361	10	F06455
66	18	2.4	369	10	W07561
67	18	2.4	370	10	BG379573
68	18	2.4	374	9	AA195532
69	18	2.4	375	10	BF429883
70	18	2.4	376	9	AI005438
71	18	2.4	380	10	T33416
72	18	2.4	381	9	AA856299
73	18	2.4	382	10	W58936
74	18	2.4	390	10	T78849
75	18	2.4	392	9	AI164059
76	18	2.4	392	10	BG909279
77	18	2.4	393	9	AA019771
78	18	2.4	397	9	BB836241
79	18	2.4	405	9	AA800869
80	18	2.4	406	10	BM235527
81	18	2.4	413	9	BB743686
82	18	2.4	414	9	AA801202
83	18	2.4	414	9	BB786837
84	18	2.4	419	9	BB679525
85	18	2.4	421	9	AA019015
86	18	2.4	422	9	AI243803
87	18	2.4	425	9	AA019018
88	18	2.4	426	9	AW124025
89	18	2.4	429	10	BE662049
90	18	2.4	432	9	AI000878

AV614200 AV614200
 BI288757 UI-R-DKO-
 AW914186 EST345490
 BE273587 601143045
 BI103654 602889123
 BE954362 UI-M-CCI-
 AA679456 ac50all.s
 AW758389 874007G10
 BF418859 UI-R-BJ2-
 AT244882 q198e05.x
 BE944264 UI-M-BH3-
 BG790803 sae70c03.
 A0759251 HS-5451.B
 BM386473 UI-R-CN1-
 EG383314 301304 MA
 BF279171 CA_EB003
 AI731923 BNLGH113
 AW141069 EST291092
 A0453959 HS-5165.B
 A2446530 1M0243002
 AZ858721 2M0164605
 B66149 C1T-HSP-202
 AZ978692 2M0255E11
 BF275075 GA_EB002
 BF275073 GA_EB002
 BI829148 603079250
 BF275344 GA_EB002
 BF363397 CH230-48F
 BI462868 603207B51
 AK013777 Mus muscu
 AA980784 ua45e12.r
 BB055152 BB055152
 AA062149 m185f04.r
 AL119042 DKE2761P
 AW312313 4011 MARC
 W36972 mb75c10.r1
 BF290469 EST455060
 AW662992 hh83h02.y
 BB491192 BB491192
 AA751893 96AS0752
 AA671815 v102407.r
 BB24526 BB24526
 AA350438 EST57699
 AA285185 zs60b06.r
 AI372485 EST175320
 AQ918301 RPCI-23-2
 W36801 mb90d01.r1
 F06455 HSC186111.n
 W07561 za98h03.r1
 BG379573 UI-R-C50-
 AA195532 mv40a05.r
 BF429883 255785 MA
 AI005438 ou08d02.x
 T33416 EST57781.Hu
 AA856299 w99h03.r
 W58936 md17e11.r1
 T78849 yc96h08.r1
 AI164059 A054P39U
 BG909279 Taur1175F
 AA019771 ze62e03.r
 BB836241 BB836241
 AA800869 EST190366
 BM235527 K0420B10-
 BB743686 BB743686
 AA801202 EST190699
 BB786837 BB786837
 BB679525 BB679525
 AA019015 ze58e08.r
 AI243803 gh84a03.x
 AA019018 ze58f08.r
 AW124025 UI-M-BH2.
 BE662049 bs16f01.y
 AI000878 os55g05.s

91	18	2.4	432	10	BF761516	164	18	2.4	554	9	AA475266	vh20b05..r
c 92	18	2.4	433	9	BF774766	c 165	18	2.4	556	10	BM380517	BM380517
c 93	18	2.4	434	9	BB670938	c 166	18	2.4	556	10	BM382283	BM382283
c 94	18	2.4	435	12	AQ087140	c 167	18	2.4	560	9	AU035070	AU035070
c 95	18	2.4	436	9	BB774982	c 168	18	2.4	563	9	AW274312	AW274312
c 96	18	2.4	439	9	A1504744	c 169	18	2.4	567	10	BM229192	BM229192
c 97	18	2.4	440	12	AZ310960	c 170	18	2.4	567	10	BM231509	BM231509
c 98	18	2.4	443	9	AU017959	c 171	18	2.4	567	10	BM233134	BM233134
c 99	18	2.4	443	9	BB776762	c 172	18	2.4	568	10	BM079055	BM079055
c 100	18	2.4	443	9	BB835101	c 173	18	2.4	570	10	BM052937	BM052937
c 101	18	2.4	443	12	AQ231724	c 174	18	2.4	571	10	HE976439	HE976439
c 102	18	2.4	444	9	BB785377	c 175	18	2.4	574	10	BM234275	BM234275
c 103	18	2.4	451	12	AQ225794	c 176	18	2.4	576	10	BM234275	BM234275
c 104	18	2.4	452	12	AQ427190	c 177	18	2.4	582	9	AA619334	AA619334
c 105	18	2.4	454	9	BB831140	c 178	18	2.4	584	9	AA619334	AA619334
c 106	18	2.4	457	9	BB787118	c 179	18	2.4	585	9	AA619334	AA619334
c 107	18	2.4	460	12	AQ231642	c 180	18	2.4	589	9	AA619334	AA619334
c 108	18	2.4	461	10	BF407691	c 181	18	2.4	591	12	AA619334	AA619334
c 109	18	2.4	462	9	A1760897	c 182	18	2.4	593	10	AA619334	AA619334
c 110	18	2.4	462	12	AQ061367	c 183	18	2.4	595	12	AA619334	AA619334
c 111	18	2.4	466	9	AA060260	c 184	18	2.4	602	10	AA619334	AA619334
c 112	18	2.4	467	9	AA613546	c 185	18	2.4	602	10	AA619334	AA619334
c 113	18	2.4	468	10	BB864033	c 186	18	2.4	603	12	AA619334	AA619334
c 114	18	2.4	469	10	BB864033	c 187	18	2.4	603	12	AA619334	AA619334
c 115	18	2.4	470	9	A1892781	c 188	18	2.4	607	10	AA619334	AA619334
c 116	18	2.4	471	9	AA065114	c 189	18	2.4	609	10	AA619334	AA619334
c 117	18	2.4	472	9	A1455323	c 190	18	2.4	610	10	AA619334	AA619334
c 118	18	2.4	472	12	CNS00RP5	c 191	18	2.4	610	10	AA619334	AA619334
c 119	18	2.4	476	9	BB820999	c 192	18	2.4	613	10	AA619334	AA619334
c 120	18	2.4	478	9	AA165322	c 193	18	2.4	613	10	AA619334	AA619334
c 121	18	2.4	480	9	A1231057	c 194	18	2.4	615	10	AA619334	AA619334
c 122	18	2.4	480	9	A1945701	c 195	18	2.4	625	10	AA619334	AA619334
c 123	18	2.4	482	10	BE979474	c 196	18	2.4	626	10	AA619334	AA619334
c 124	18	2.4	483	9	A1946868	c 197	18	2.4	632	9	AA619334	AA619334
c 125	18	2.4	484	9	A1451565	c 198	18	2.4	635	9	AA619334	AA619334
c 126	18	2.4	486	9	AA305051	c 199	18	2.4	635	10	AA619334	AA619334
c 127	18	2.4	489	10	BE287534	c 200	18	2.4	635	10	AA619334	AA619334
c 128	18	2.4	490	12	AZ225229	c 201	18	2.4	638	10	AA619334	AA619334
c 129	18	2.4	492	10	BM380051	c 202	18	2.4	640	10	AA619334	AA619334
c 130	18	2.4	494	12	AQ2897915	c 203	18	2.4	642	12	AA619334	AA619334
c 131	18	2.4	503	9	A1945708	c 204	18	2.4	648	9	AA619334	AA619334
c 132	18	2.4	506	10	BG637771	c 205	18	2.4	652	10	AA619334	AA619334
c 133	18	2.4	509	9	AW140477	c 206	18	2.4	655	10	AA619334	AA619334
c 134	18	2.4	509	10	BM248576	c 207	18	2.4	661	9	AA619334	AA619334
c 135	18	2.4	510	10	BG634152	c 208	18	2.4	661	10	AA619334	AA619334
c 136	18	2.4	510	10	BE976470	c 209	18	2.4	663	10	AA619334	AA619334
c 137	18	2.4	512	9	A1383513	c 210	18	2.4	664	10	AA619334	AA619334
c 138	18	2.4	513	9	A1418838	c 211	18	2.4	667	10	AA619334	AA619334
c 139	18	2.4	513	9	A1760811	c 212	18	2.4	669	9	AA619334	AA619334
c 140	18	2.4	514	10	BE372378	c 213	18	2.4	672	10	AA619334	AA619334
c 141	18	2.4	517	10	BF192481	c 214	18	2.4	673	9	AA619334	AA619334
c 142	18	2.4	518	9	BF755077	c 215	18	2.4	673	10	AA619334	AA619334
c 143	18	2.4	519	10	A1012485	c 216	18	2.4	675	10	AA619334	AA619334
c 144	18	2.4	523	9	BF758598	c 217	18	2.4	681	9	AA619334	AA619334
c 145	18	2.4	524	10	B1548713	c 218	18	2.4	682	9	AA619334	AA619334
c 146	18	2.4	525	10	R60284	c 219	18	2.4	682	12	AA619334	AA619334
c 147	18	2.4	526	9	AA003909	c 220	18	2.4	685	10	AA619334	AA619334
c 148	18	2.4	526	9	AA527134	c 221	18	2.4	690	10	AA619334	AA619334
c 149	18	2.4	526	10	W17319	c 222	18	2.4	691	9	AA619334	AA619334
c 150	18	2.4	529	9	AW367279	c 223	18	2.4	692	10	AA619334	AA619334
c 151	18	2.4	531	9	BE230327	c 224	18	2.4	701	9	AA619334	AA619334
c 152	18	2.4	531	12	AQ208135	c 225	18	2.4	702	12	AA619334	AA619334
c 153	18	2.4	533	10	BF372376	c 226	18	2.4	704	10	AA619334	AA619334
c 154	18	2.4	534	10	BF494649	c 227	18	2.4	708	10	AA619334	AA619334
c 155	18	2.4	535	9	BB769449	c 228	18	2.4	710	10	AA619334	AA619334
c 156	18	2.4	537	9	AA637673	c 229	18	2.4	721	10	AA619334	AA619334
c 157	18	2.4	539	9	BB831407	c 230	18	2.4	726	9	AA619334	AA619334
c 158	18	2.4	539	10	BG634774	c 231	18	2.4	739	10	AA619334	AA619334
c 159	18	2.4	541	10	BM229152	c 232	18	2.4	741	10	AA619334	AA619334
c 160	18	2.4	544	9	AW540478	c 233	18	2.4	743	12	AA619334	AA619334
c 161	18	2.4	549	12	AZ222239	c 234	18	2.4	745	10	AA619334	AA619334
c 162	18	2.4	551	9	BB752007	c 235	18	2.4	795	9	AA619334	AA619334
c 163	18	2.4	553	9	AV608039	c 236	18	2.4	795	12	AA619334	AA619334

```

237 18 2.4 826 12 AQ050438 nbxb0004a
c 238 18 2.4 833 10 W28239 4422 Human
c 239 18 2.4 833 10 BF214011
240 18 2.4 844 10 BF754100
241 18 2.4 846 10 BG071028
242 18 2.4 846 12 BH293013
c 243 18 2.4 861 12 AQ741500
c 244 18 2.4 867 10 B1413457
245 18 2.4 873 10 BE619378
c 246 18 2.4 889 12 CNS010KE
c 247 18 2.4 896 12 CNS020NY
c 248 18 2.4 906 10 BF181302
c 249 18 2.4 910 10 B1415323
250 18 2.4 921 12 B19154

```

ALIGNMENTS

```

RESULT 1
A1654622/c
LOCUS
DEFINITION wb48f11.x1 NCI_CGAP_GC6 Homo sapiens CDNA clone IMAGE:2308941 3',
EST.
ACCESSION A1654622.1 GI:4738601
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 472 Std Error: 0.00
Seq primer: -40UP from Gibco.

```

```

FEATURES
source
1..384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2308941"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 81 a 72 c 119 g 112 t
ORIGIN

```

```

Query Match 16.6%; Score 124; DB 9; Length 384;
Best Local Similarity 100.0%; Pred. No. 8.1e-55;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 623 GTATCTCTCTCTCCAGAGAGATATGTCCACTGATCCGATCTTTCATCCATCCCTG 682
Db 384 GTATCTCTCTCTCCAGAGAGATATGTCCACTGATCCGATCTTTCATCCATCCCTG 325
QY 683 GATGTGAATGTCTCAGCAGGTCGCGATATCTCAGCAGCAAAACCCAGCTGCAAAACC 742
Db 324 GATGTGAATGTCTCAGCAGGTCGCGATATCTCAGCAGCAAAACCCAGCTGCAAAACC 265
QY 743 CTCGA 746
Db 264 CTCGA 261

```

```

RESULT 2
BE074387/c
LOCUS
DEFINITION QV3-BT0572-030200-079-d03 BT0572 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE074387
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 117)
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nazari, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bata, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LIUR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV3-BT0572-030)
200-079-d03et3=2000-02-03et4=1)
Seq primer: puc 18 forward
High quality sequence stop: 117.

```

```

FEATURES
source
1..117
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="BT0572"
/dev_stage="Adult"
/notes="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 23 a 29 c 35 g 30 t
ORIGIN

```

```

Query Match 6.7%; Score 50; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.5e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 697 TCAGCAGGTCTCGATATCTCTCAGCAGCAAAACCCAGCTGCAAAACCTCA 746
|||||

```



```

Db 117 TCAGCAGGTGCTCGATAATCTCAGCACAAAACCCAGCTGCAAAACCTCA 68

RESULT 3
BE074388/c
LOCUS
DEFINITION QV3-BT0572-030200-079-e01 BT0572 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE074388
VERSION BE074388.1 GI:8422259
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 117)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=8t2-QV3-BT0572-030
200-079-e01&t3=2000-02-03&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 117.
FEATURES
Location/Qualifiers
1..117
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0572"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 22 a 29 c 36 g 30 t
ORIGIN

Query Match 6.7%; Score 50; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.5e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 TCAGCAGGTGCTCGATAATCTCAGCACAAAACCCAGCTGCAAAACCTCA 746
|||||
Db 117 TCAGCAGGTGCTCGATAATCTCAGCACAAAACCCAGCTGCAAAACCTCA 69

RESULT 5
BE074390/c
LOCUS
DEFINITION QV3-BT0572-030200-079-g01 BT0572 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE074390
VERSION BE074390.1 GI:8422263
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 130)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=8t2-QV3-BT0572-030
200-079-e01&t3=2000-02-03&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 118.
FEATURES
Location/Qualifiers
1..118
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0572"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 22 a 29 c 37 g 30 t
ORIGIN

Query Match 6.7%; Score 50; DB 9; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.5e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 TCAGCAGGTGCTCGATAATCTCAGCACAAAACCCAGCTGCAAAACCTCA 746
|||||
Db 118 TCAGCAGGTGCTCGATAATCTCAGCACAAAACCCAGCTGCAAAACCTCA 69

RESULT 4
BE074383/c
LOCUS
DEFINITION QV3-BT0572-030200-079-c01 BT0572 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE074383
VERSION BE074383.1 GI:8422251
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=8t2-QV3-BT0572-030
200-079-e01&t3=2000-02-03&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 117.
FEATURES
Location/Qualifiers
1..117
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0572"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 22 a 29 c 36 g 30 t
ORIGIN

Query Match 6.7%; Score 50; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.5e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 TCAGCAGGTGCTCGATAATCTCAGCACAAAACCCAGCTGCAAAACCTCA 746
|||||
Db 117 TCAGCAGGTGCTCGATAATCTCAGCACAAAACCCAGCTGCAAAACCTCA 68

```

JOURNAL
MEDLINE
COMMENT

sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-BT0572-030
200-079-q01st3-2000-02-03st4-1)
Seq primer: puc 18 forward
High quality sequence stop: 130.
Location/Qualifiers
1. .130

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0572"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
26 a 31 c 40 g 33 t

BASE COUNT

ORIGIN

Query Match 6.6% Score 49; DB 9; Length 130;
Best Local Similarity 100.0%; Pred. No. 5.2e-15;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6

BE074380/c
LOCUS BE074380 117 bp mRNA linear EST 09-JUN-2000
DEFINITION QV3-BT0572-030200-079-a08 BT0572 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE074380
VERSION BE074380.1 GI:8422245
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 117)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

TITLE

JOURNAL
PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-BT0572-030
200-079-q08st3-2000-02-03st4-1)
Seq primer: puc 18 forward
High quality sequence stop: 117.
Location/Qualifiers
1. .117

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0572"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
22 a 30 c 35 g 30 t

BASE COUNT

ORIGIN

Query Match 5.1% Score 38; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.6e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 709 CGATAATCCTCAGCACAAACCCAGCTGCAAAACCTCA 746

Db 105 CGATAATCCTCAGCACAAACCCAGCTGCAAAACCTCA 68

RESULT 7

BE074379
LOCUS BE074379 119 bp mRNA linear EST 09-JUN-2000
DEFINITION QV3-BT0572-030200-079-a05 BT0572 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE074379
VERSION BE074379.1 GI:8422243
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 119)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

TITLE

JOURNAL
PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-BT0572-030
200-079-a05st3-2000-02-03st4-1)
Seq primer: puc 18 forward
High quality sequence stop: 119.
Location/Qualifiers
1. .119

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone_lib="BT0572"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 30 a 36 c 31 g 22 t
ORIGIN

Query Match 4.8%; Score 36; DB 9; Length 119;
Best Local Similarity 100.0%; Pred. No. 4.3e-08;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 711 ATATCTCTCAGCACAAACCCAGCTGCAAAACCTCA 746
|||||
Db 15 ATATCTCTCAGCACAAACCCAGCTGCAAAACCTCA 50

RESULT 8

LOCUS BE074382 119 bp mRNA linear EST 09-JUN-2000
DEFINITION QV3-BT0572-030200-079-b06 BT0572 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE074382
VERSION BE074382.1 GI:8422249
KEYWORDS EST.
SOURCE Homo sapiens

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 119)

REFERENCE
AUTHORS Dias Neto, B., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICK Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-BT0572-030200-079-b06&t3=2000-02-03&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 119.

Location/Qualifiers
1..119

FEATURES
Source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0572"
/dev_stage="Adult"

/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 29 a 36 c 32 g 22 t
ORIGIN

Query Match 4.4%; Score 33; DB 9; Length 119;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 714 ATCTCTCAGCACAAACCCAGCTGCAAAACCTCA 746
|||||
Db 18 ATCTCTCAGCACAAACCCAGCTGCAAAACCTCA 50

RESULT 9

LOCUS BF651530 406 bp mRNA linear EST 25-APR-2001
DEFINITION 274330 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF651530
VERSION BF651530.1 GI:11916660
KEYWORDS EST.
SOURCE cow.

ORGANISM
Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 406)

REFERENCE
AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR primers
FORWARD: AGGAACACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGAGC

Plate: 63 row: P column: 17
Seq primer: ATTTAGGTGACACTATAG.

Location/Qualifiers
1..406

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

BASE COUNT 103 a 85 c 110 g 108 t
ORIGIN

FEATURES
source

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

Query Match 3.1%; Score 23; DB 10; Length 406;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TTCAGCTTTGGAACTGTGTCTC 28
|||||
Db 45 TTCAGCTTTGGAACTGTGTCTC 67

RESULT 10
AZ320113

LOCUS AZ320113 270 bp DNA linear GSS 29-SEP-2000
 DEFINITION 1M0040D04F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
 clone UUGCLM0040D04 F, DNA sequence.
 ACCESSION AZ320113
 VERSION AZ320113.1 GI:10371565
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 270)
 REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D.,Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0040 row: D column: 04
 Seq primer: CGTTGTAAACGACGCCAGT
 Class: plasmid ends
 High quality sequence stop: 270.
 Location/Qualifiers
 1..270
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGCLM0040D04"
 /clone_lib="Mouse 10kb plasmid UUGCLM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (q1147321141gb1A129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

FEATURES

source

BASE COUNT 63 a 71 c 62 g 74 t
 ORIGIN

Query Match 2.8%; Score 21; DB 12; Length 270;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTTACGCTTTGGAACTTGT 25
 |||||

Db 173 CTTACGCTTTGGAACTTGT 193
 |||||

RESULT 11

LOCUS

AZ977289

DEFINITION

2M0253G12F Mouse 10kb plasmid UUGC2M library Mus musculus genomic

clone UUGC2M0253G12 F, DNA sequence.

ACCESSION

AZ977289

VERSION

AZ977289.1 GI:13848516

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 425)

REFERENCE

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D.,Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0253 row: G column: 12

Seq primer: CGTTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 425.

Location/Qualifiers

1..425

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0253G12"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (q1147321141gb1A129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 101 a 109 c 93 g 122 t
 ORIGIN

Query Match 2.8%; Score 21; DB 12; Length 425;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTTACGCTTTGGAACTTGT 25
 |||||

Db 336 CTTACGCTTTGGAACTTGT 356
 |||||

Y., Cardenas, M., McCann, R., Waterson, R., Wilson, R. and Sibley, D.
WashU-Merck Eimeria tenella project
Unpublished (1999)
Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Contact David Sibley (toxotes@orcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Trace considered overall poor quality
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
Features
Source
1..358
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria M5-6 Merozoite stage subtracted"
/dev_stage="Merozoite"
/lab_host="SOLR E. coli"
/note="vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Merozoites were obtained from ceacal scrapings of
chickens infected with E. tenella. cDNA was synthesized
from poly mRNA using an oligo-dT primer containing a XhoI
site. Following second strand synthesis, EcoRI adapters
were ligated to the cDNA and products were size-selected
on Sephacryl S500. The cDNAs were ligated to EcoRI/XhoI
prepared lambda ZAP11(Stratagene). Clones were converted
to phagemids by mass excision using EXAssist helper phage
and E.coli SOLR cells (Stratagene). Insert sizes range
from 0.7-1.5Kb. The library may contain a small percentage
of host or bacterial contaminants. Clones were selected by
negative hybridization against a pool of over-represented
ESTs (N>=10, from 1506 previous reads)."
BASE COUNT 102 a 67 c 91 g 97 t 1 others
ORIGIN

Query Match 2.7%; Score 20; DB 9; Length 358;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 TTCTTCACGAGGGACTTCGAG 136
|||||
DB 246 TTCTTCACGAGGGACTTCGAG 227

RESULT 14
BE028127/c
LOCUS
DEFINITION
Eimeria M5-6 Merozoite stage subtracted Eimeria
tenella cDNA 5', similar to SW:RL22_TRIGR PI3732 60S RIBOSOMAL
PROTEIN L22 ;, mRNA sequence.

ACCESSION BE028127
VERSION BE028127.1 GI:8321542
KEYWORDS EST.
SOURCE Eimeria tenella
ORGANISM Eimeria tenella
Eukaryote; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
Eimeria.
1 (bases 1 to 456)
Liberator, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,
Martin, J., Wylie, T., Underwood, K., Seftoe, M., Theising, B., Allen
M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey
N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson
Y., Cardenas, M., McCann, R., Waterson, R., Wilson, R. and Sibley, D.
WashU-Merck Eimeria tenella project
Unpublished (1999)
Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Contact David Sibley (toxost@borcim.wustl.edu) for further
 information relating to organism, libraries, or clone availability.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 427.
 Location/Qualifiers

FEATURES

source
 1..456
 /organism="Eimeria tenella"
 /strain="IS18"
 /db_xref="taxon:5802"
 /clone_lib="Eimeria M5-6 Merozoite stage subtracted"
 /dev_stage="Merozoite"
 /lab_host="SOLR E. coli"
 /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
 ; Merozoites were obtained from ceecal scrapings of
 chickens infected with E. tenella. cDNA was synthesized
 from poly mRNA using an oligo-dT primer containing a XhoI
 site. Following second strand synthesis, EcoRI adapters
 were ligated to the cDNA and products were size-selected
 on Sephacryl S500. The cDNAs were ligated to EcoRI/XhoI
 prepared lambda ZAPII(Stratagene). Clones were converted
 to phagemids by mass excision using Exassist helper phage
 and E.Coli SOLR Cells (Stratagene). Insert sizes range
 from 0.7-1.5Kb. The library may contain a small percentage
 of host or bacterial contaminants. Clones were selected by
 negative hybridization against a pool of over-represented
 ESTs (N>=10, from 1506 previous reads)."
 BASE COUNT 112 a 102 c 124 g 116 t 2 others
 ORIGIN

Query Match 2.7%; Score 20; DB 9; Length 456;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 TTCTTCACGAGGACTTCAG 136
 |||||
 DB 362 TTCTTCACGAGGACTTCAG 343

RESULT 15

LOCUS BE751494
 DEFINITION 203654 MARC 2BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
 ACCESSION BE751494
 VERSION BE751494.1 GI:10165486
 KEYWORDS EST.
 SOURCE COW.

ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 475)

REFERENCE Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
 G.L., Heaton,W.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perle,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE 21180013
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred

FEATURES

source
 1..553
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 2BOV"
 /tissue_type="pooled"

v0.980904.e. vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCAGTCACGACG
 Plate: 43 row: F column: 19
 Seq primer: ATTTAGTGACACTATAG.

FEATURES

source
 1..475
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 2BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."
 BASE COUNT 93 a 162 c 138 g 82 t
 ORIGIN

Query Match 2.7%; Score 20; DB 10; Length 475;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 ATCTTCATCCACTCCCTGGA 584
 |||||
 DB 209 ATCTTCATCCACTCCCTGGA 228

RESULT 16

LOCUS BE751485
 DEFINITION 203644 MARC 2BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
 ACCESSION BE751485
 VERSION BE751485.1 GI:10165477
 KEYWORDS EST.
 SOURCE COW.

ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 553)

REFERENCE Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
 G.L., Heaton,W.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perle,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE 21180013
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

TITLE

JOURNAL USDA, ARS, US Meat Animal Research Center
 MEDLINE PO Box 166, Clay Center, NE 68933-0166, USA
 COMMENT Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCAGTCACGACG
 Plate: 43 row: D column: 21
 Seq primer: ATTTAGTGACACTATAG.

Location/Qualifiers
 1..553
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 2BOV"
 /tissue_type="pooled"

```

/lab_host="DH109"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT      122 a   178 c   159 g   94 t
ORIGIN

Query Match      2.7%; Score 20; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 ATCTTCATCCACCTCCCTGGA 684
      |||||||
Db 209 ATCTTCATCCACCTCCCTGGA 228

RESULT 17
LOCUS      B1540670          574 bp mRNA linear EST 30-AUG-2001
DEFINITION 454023 MARC lBOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  B1540670
VERSION    B1540670.1 GI:15381782
KEYWORDS   EST.
SOURCE     COW.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1 (bases 1 to 574)
AUTHORS    Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrkrug,S.C., Bennett
            G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
            Perle,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
TITLE      Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL    Genome Res. 11 (4), 626-630 (2001)
MEDLINE    21180013
COMMENT    Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAG
Plate: 118 row: B column: 6
Seq primer: ATTAGGTGACACTATAG.
            Location/Qualifiers
                1..574
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone_lib="MARC lBOV"
                /tissue_type="pooled"
                /lab_host="DH109"
                /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                Library made from pooled tissue from lymph node, ovary,
                fat, hypothalamus, and pituitary."
BASE COUNT      143 a   124 c   208 g   98 t   1 others
ORIGIN

Query Match      2.7%; Score 20; DB 10; Length 574;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 AACGACAGGCCCGGAGGC 243
      |||||||
Db 222 AACGACAGGCCCGGAGGC 241

RESULT 19
LOCUS      BI288757/c          661 bp mRNA linear EST 19-JUL-2001
DEFINITION UI-R-DKO-cde-e-11-0-UI.s1 UI-R-DKO Rattus norvegicus cDNA clone
ACCESSION  BI288757
VERSION    BI288757.1 GI:14945645
KEYWORDS   EST.
SOURCE     Norway rat.
            Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 661)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene

```

```

|||||
Db 259 AACGACAGGCCCGGAGGC 278

RESULT 18
LOCUS      AV614200          578 bp mRNA linear EST 28-NOV-2001
DEFINITION AV614200 Bos taurus adipocyte cell line Bos taurus cDNA clone
ACCESSION  E1AD003F02 5', mRNA sequence.
VERSION    AV614200
KEYWORDS   EST.
SOURCE     COW.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1 (bases 1 to 578)
AUTHORS    Takasuga,A., Hirotsune,S., Itoh,R., Jitchazono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
TITLE      Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
JOURNAL    Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE    21570554
COMMENT    Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shikakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@ccoc.oca.nie.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
            Location/Qualifiers
                1..578
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone_lib="E1AD003F02"
                /cell_type="an adipocyte cell line"
                /lab_host="DH109"
                /note="Vector: p2L1; Site_1: Sail1; Site_2: NotI; Poly A
                was deleted from a NotI site"
BASE COUNT      138 a   133 c   209 g   97 t   1 others
ORIGIN

Query Match      2.7%; Score 20; DB 9; Length 578;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 AACGACAGGCCCGGAGGC 243
      |||||||
Db 222 AACGACAGGCCCGGAGGC 241

RESULT 19
LOCUS      BI288757/c          661 bp mRNA linear EST 19-JUL-2001
DEFINITION UI-R-DKO-cde-e-11-0-UI.s1 UI-R-DKO Rattus norvegicus cDNA clone
ACCESSION  BI288757
VERSION    BI288757.1 GI:14945645
KEYWORDS   EST.
SOURCE     Norway rat.
            Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 661)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene

```

Wed Aug 7 05:46:20 2002

discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat kidney pool library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers

1..661
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DK0-cde-e-11-0-UI"
/clone_lib="UI-R-DK0"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker: Site_1: Not I; Site_2: Eco RI; The UI-R-DK0 library is a subtracted library derived from a mixture of five individually tagged normalized rat libraries: brain-nRBP (20%), heart-nRBP (20%), kidney-nRBP (20%), aorta-nRBP (20%), and placenta-nRBP (20%). Each original library was constructed from a mixture of equal amounts of RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, and adult day 200. (Exception: the aorta pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research genome 6: 791-806, 1996). For construction of the DK0 subtracted library, plasmid DNA from each of the five individually tagged normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded circular DNA representing these five normalized libraries was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a set of about 1,000 arrayed clones from each of the five non-normalized libraries of brain (CTOs), heart (CSOs), kidney (CUOs), aorta (CWUs), and placenta (CXOs). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (CTOs), heart (CSOs), kidney (CUO), aorta (CWU), and placenta (CXO). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population.

TAG_LiB=UI-R-DK0
TAG_TISSUE=rat kidney pool
TAG_SEQ=CAAGACTGTC
144 a 192 c 185 g 138 t 2 others
2.7%; Score 20; DB 10; Length 661;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT

144 a 192 c 185 g 138 t 2 others

Query Match

Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 CAGTTGACATACTCTTAAA 157
|||||
DB 35 CAGTTGACATACTCTTAAA 16

RESULT 20

AW914186/c 690 bp mRNA linear EST 25-MAY-2000
LOCUS EST345490 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
DEFINITION RGIAB55 5' end, mRNA sequence.

AW914186

ACCESSION AW914186.1 GI:8079860

VERSION EST.

KEYWORDS Rattus sp.

SOURCE Rattus sp.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 690)

AUTHORS Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,

Kerlavage, A.R. and Adams, M.D.

Rat Genome Project; Generation of a Rat EST (REST) Catalog & Rat

Gene Index

Unpublished (1998)

CONTACT: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

This clone is available through the ATCC, contact the ATCC

tel#703-365-2700 for further information

Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

1..690

Source

/organism="Rattus sp."

/db_xref="taxon:10118"

/clone="RGIAB55"

/clone_lib="Normalized rat brain, Bento Soares"

/note="Organ: brain; Vector: pf73Pac; Site_1: EcoRI;

Site_2: NotI"

155 a 183 c 184 g 168 t

BASE COUNT

155 a 183 c 184 g 168 t

ORIGIN

Query Match 2.7%; Score 20; DB 9; Length 690;

Best Local Similarity 100.0%; Pred. No. 24;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 GGACCTCAGTCCTCTCT 66

|||||

DB 381 GGACCTCAGTCCTCTCT 362

RESULT 21

BE273587/c 695 bp mRNA linear EST 26-OCT-2000
LOCUS 601143045F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:3506795 5',
DEFINITION mRNA sequence.

BE273587

ACCESSION BE273587.1 GI:9148375

VERSION EST.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 695)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Site_2: SalI: Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI-CCAP Library. I"

BASE COUNT 259 a 249 c 220 g 151 t

Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc
Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI85 row: J column: 12
High quality sequence stop: 534.

FEATURES
source

Location/Qualifiers
1..695

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3506795"
/clone_lib="NHL-CCAP-14"
/issue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 191 a 118 c 220 g 166 t

Query Match 2.7%; Score 20; DB 10; Length 595;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 AACCTGAGCTTCCTGTCCAC 390
|||||

DB 565 AACCTGAGCTTCCTGTCCAC 646

RESULT 22

LOCUS B1103654 879 bp mRNA linear EST 26-JUN-2001
DEFINITION 602889133F1 NCI-CCAP_Kid14 Mus musculus cDNA clone IMAGE:5044421
5', mRNA sequence.

ACCESSION B1103654

VERSION B1103654.1 GI:14554547

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: scapsh-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14M1121 row: n column: 06

High quality sequence stop: 542.

Location/Qualifiers

1..879

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5044421"

/clone_lib="NCI-CCAP_Kid14"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: kidney; Vector: PCMV-SPOK76; Site_1: NotI;

FEATURES
source

BASE COUNT 88 a 143 c 99 g 85 t

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 71;

Query Match 2.7%; Score 20; DB 10; Length 879;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 517 AGAATGCCGACGTGACCCAA 536

DB 734 AGAATGCCGACGTGACCCAA 753

RESULT 23

BE954362

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Bonaldi,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

MEDLINE

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mes@mail.nih.gov

Oligo-dT track not found, Not I site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab Clone distribution: Researchers may obtain BMAP cDNA

clones from RESEARCH GENETICS. It should be noted that Bento Soares

is generating a small number of additional specialized

non-redundant arrays of BMAP cDNAs whose availability will be

considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1..415

/organism="Mus musculus"

/strain="C57Bl/6J"

/db_xref="taxon:10090"

/clone="01-M-CC1-bae-d-01-0-UI"

/clone_lib="NIH-BMAP_Ret1_N"

/dev_stage="13.5 days pc"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The

NIH-BMAP_Ret1_N library is a normalized library derived

NIH-BMAP_Ret1_N library. For a detailed description of the

embryonic retina tissue. For a detailed description of the

library from which this clone was derived, please visit

our web site at brainest.eng.uiowa.edu.

TAG_SEQ=None found"

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 626 TCTCCCTGCTGCAGAGG 644
|||||
Db 71 TCTCCCTGCTGCAGAGG 89

RESULT 24
AA679456 421 bp mRNA linear EST 02-DEC-1997
LOCUS ac50a11.s1 Stratagene hMT neuron (#937233) Homo sapiens cDNA clone
DEFINITION IMAGE:859868 3' similar to contains Alu repetitive element, mRNA
sequence.

ACCESSION AA679456
VERSION AA679456
KEYWORDS AA679456.1 GI:2659978
SOURCE EST.
ORGANISM human.

REFERENCE 1 (bases 1 to 421)
AUTHORS Hillier,D., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 411.

FEATURES
source
1..421
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:859868"
/dev_stage="hMT neurons"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Differentiated, post mitotic hMT neurons. Average insert size: 1.5 kb; Uni-ZAP XR vector: -5' adaptor sequence: 5' GAATTCGGCCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 92 a 111 c 109 g 109 t

ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 367 CCTTAACCTGAGCTTCCT 385
|||||
Db 356 CCTTAACCTGAGCTTCCT 374

RESULT 25
AW758389 424 bp mRNA linear EST 03-MAY-2000
LOCUS 874007G10.xl C. reinhardtii CC-1690, Lambda Zap II Chlamydomonas
DEFINITION reinhardtii cDNA, mRNA sequence.

ACCESSION AW758389
VERSION AW758389.1 GI:7687743
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii
Eukaryota: Viridiplantae: Chlorophyta: Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 424)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants: project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Elizabeth H. Harris
DCMB Box 91000
Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.

FEATURES
source
1..424
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the Lambda Zap clones by superinfection with ExAssist (Stratagene) phage."

BASE COUNT 102 a 136 c 95 g 91 t

ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 424;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 397 TGTCACGTGCGCGGCC 415
|||||
Db 260 TGTCACGTGCGCGGCC 278

RESULT 26
BF418859 472 bp mRNA linear EST 28-NOV-2000
LOCUS UI-R-RJ2-bqj-g-06-0-U1.s1 UI-R-RJ2 Rattus norvegicus cDNA clone
DEFINITION UI-R-RJ2-bqj-g-06-0-U1 3', mRNA sequence.

ACCESSION BF418859
VERSION BF418859.1 GI:11406848
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota: Metazoa: Chordata: Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 472)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical
Tel: 319 335 8250
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized AV canal at 16.5 dpc library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..472
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bqj-g-06-0-UI"
/clone_lib="UI-R-BJ2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-BJ2
TAG_TISSUE=AV canal at 16.5 dpc
TAG_SEQ=GAACC"

BASE COUNT 116 a 105 C 93 g 158 t
ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CAGAGGCCCGAGGAGCTG 245
|||||
DB 118 CAGAGGCCCGAGGAGCTG 136

RESULT 27
A1244882/C
LOCUS A1244882 477 bp mRNA linear EST 28-JAN-1999
DEFINITION qj38e05.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1867520 3'; mRNA sequence.
ACCESSION A1244882
VERSION A1244882.1 GI:3840279
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: caapbs@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 721 Std Error: 0.00

Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized AV canal at 16.5 dpc library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..472
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bqj-g-06-0-UI"
/clone_lib="UI-R-BJ2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-BJ2
TAG_TISSUE=AV canal at 16.5 dpc
TAG_SEQ=GAACC"

BASE COUNT 116 a 105 C 93 g 158 t
ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 472;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CAGAGGCCCGAGGAGCTG 245
|||||
DB 118 CAGAGGCCCGAGGAGCTG 136

RESULT 27
A1244882/C
LOCUS A1244882 477 bp mRNA linear EST 28-JAN-1999
DEFINITION qj38e05.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1867520 3'; mRNA sequence.
ACCESSION A1244882
VERSION A1244882.1 GI:3840279
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: caapbs@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 721 Std Error: 0.00

Seq primer: -40UP from Gibco
High quality sequence stop: 422.

FEATURES
source
Location/Qualifiers
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1867520"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 145 a 98 C 80 g 154 t
ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 714 ATCTCAGCACAAACCCA 732
|||||
DB 419 ATCTCAGCACAAACCCA 401

RESULT 28
BE944264
LOCUS BE944264 478 bp mRNA linear EST 03-OCT-2000
DEFINITION UI-M-BH3-avt-a-07-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone.
ACCESSION BE944264
VERSION BE944264.1 GI:10522023
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 478)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.

FEATURES
source
Location/Qualifiers
1..478
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-avt-a-07-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; The
 NIH_BMAP_M_S4 library is a subtracted library of a series,
 ultimately derived from a mixture of individually tagged
 normalized libraries from ten regions of the mouse brain
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,
 cortex, amygdala, basal ganglia, pineal gland, striatum,
 hippocampus) after a series of subtractions to reduce the
 representation of cDNAs from which ESTs had already been
 generated. The following serially subtracted libraries
 were generated in this process: NIH_BMAP_M_S4,
 NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
 NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
 (NIH_BMAP_M_S4) was constructed as follows: PCR amplified
 cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
 NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
 was used as a driver in a hybridization with a pool of
 the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
 libraries in the form of single-stranded circles. The
 remaining single-stranded circles (subtracted library)
 was purified by hydroxyapatite column chromatography,
 converted to double-stranded circles and electroporated
 into DH10B bacteria (Life Technologies) to generate the
 NIH_BMAP_M_S4 library. This procedure has been previously
 described (Bonaldo, Lennon and Soares, Genome Research
 6:791-806, 1996)
 TAG_SEQ=None found"
 BASE COUNT 100 a 164 c 113 g 100 t 1 others
 ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 478;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 626 TCCTCCCTGCTGCAGAGG 644
 |||||
 Db 71 TCCTCCCTGCTGCAGAGG 89

RESULT 29
 BG790803 481 bp mRNA linear EST 29-NOV-2001
 LOCUS
 DEFINITION
 sae70c03.y1 Gm-cl064 glycine max cDNA clone GENOME SYSTEMS CLONE
 ID: Gm-cl064-3558 5', similar to TR:Q9SEK2 Q9SEK2 CHLOROPLAST OUTER
 ENVELOPE HEXOKINASE 1. ;, mRNA sequence.

ACCESSION BG790803.1 GI:14126365
 VERSION
 KEYWORDS soybean.
 SOURCE
 ORGANISM
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 481)
 Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
 ,A., Bolla,B., Marita,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 ,R., Waterston,R. and Wilson,R.
 Public Soybean EST Project

TITLE
 JOURNAL
 COMMENT
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact via Email: ccure@resgen.com
 High quality sequence stop: 409.

FEATURES
 source
 1..481
 /organism="Glycine max"
 /cultivar="Williams"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl064-3558"
 /clone_lib="Gm-cl064"
 /tissue_type="seedling epicotyls"
 /dev_stage="2 week old"
 /lab_host="DH10B"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from mRNA isolated
 from the epicotyls of 2 week old seedling for the cultivar
 Williams. The seedlings were germinated in a growth
 chamber, excised above the soil level, and the plants
 were placed in a 100 ppm solution of auxin for 24 hours
 prior to harvesting. Complementary DNA was synthesized
 from mRNA using a primer consisting of a poly(dT)
 sequence with a XhoI restriction site. EcoRI adapters
 were ligated to the blunt-ended cDNA fragments followed
 by XhoI digestion. The cDNA fragments were directionally
 cloned into the EcoRI-XhoI restriction site of the
 pBluescript vector. The ligated cDNA fragments were
 transformed into DH10B host cells (GibcoBRL). This
 library was constructed in the laboratory of Dr. Randy
 Shoemaker."

BASE COUNT 126 a 80 c 137 g 138 t
 ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 481;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 152 CTTAAAGCATCCTTGAGA 170
 |||||
 Db 113 CTTAAAGCATCCTTGAGA 131

RESULT 30
 AQ799251 517 bp DNA linear GSS 09-AUG-1999
 LOCUS
 DEFINITION
 HS_5451_B1_D11_77A RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=1027 Col=21 Row=H, DNA sequence.

ACCESSION AQ799251.1 GI:5716583
 VERSION
 KEYWORDS GSS.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 517)

REFERENCE
 AUTHORS
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589

TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

or from Research Genetics (info@resgen.com). BAC end web Server:
<http://www.htsc.washington.edu>
 Plate: 1027 row: H column: 21
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 517.

FEATURES

Location/Qualifiers
 1. 517
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-1027 Col=21 Row=H"
 /clone.lib="RPGI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBAC3.6 vector at EcoRI sites"
 118 a 115 c 97 g 184 t 3 others

BASE COUNT

ORIGIN

Query Match 2.5%; Score 19; DB 12; Length 517;
 Best local Similarity 100.0%; Pred. No. 76;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AGTCTCTCTTCACAACTCT 75

Db 321 AGTCTCTCTTCACAACTCT 339

RESULT 31

BM386473/c 528 bp mRNA linear EST 17-JAN-2002
 UI-R-CNI-cjg-g-07-0-UI-s1 UI-R-CNI Rattus norvegicus cDNA clone
 UI-R-CNI-cjg-g-07-0-UI 3', mRNA sequence.

ACCESSION BM386473

VERSION BM386473.1 GI:18186526

SOURCE EST.

ORGANISM Norway rat.

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 528)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, Mj

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized brown adipose library cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

FEATURES

source

1. 528

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CNI-cjg-g-07-0-UI"

/clone.lib="UI-R-CNI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CNI
 library is a subtracted library derived from the following
 pool of seven normalized rat libraries: normalized rat
 seminal vesicles, normalized rat penis, normalized rat
 bladder, normalized rat cervix, and normalized rat
 adipose, normalized rat fundus, and normalized rat
 salivary gland. It was constructed according to the
 procedure described by Bonaldo, Lennon & Soares (Genome
 Research Genome 6: 791-806, 1996). For construction of
 the CNI library, plasmid DNA from the pool of normalized
 libraries was electroporated into competent bacteria for
 the production of single-stranded circular DNA. This was
 then used as a tracer in a subtractive hybridization with
 a driver (PCR amplified inserts from a plasmid DNA template
 preparation) comprising: a) a pool of about 34,000 clones
 from the Rat Unigene Set corresponding to plates R-5-AA-NN
 excluding plates R-5-MM and MN. This pool represented 40%
 of the final driver population. b) a pool of about 29,000
 clones from subtracted libraries CA0 and CA1 corresponding
 to plates R-CA0-AMV through R-CA0-AXS, R-CA0-AXZ through
 R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS,
 R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through
 R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA,
 R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through
 R-CA0-BOJ, R-CA0-BPA through R-CA0-BPG, R-CA0-BBA through
 R-CA0-BDA, R-CA0-BHZ through R-CA0-BJF, R-CA0-BJR,
 R-CA0-BJT through R-CA0-BKB, R-CA0-BKD, R-CA0-BKF,
 R-CA0-BKI, R-CA0-BKT, R-CA0-BLF, R-CA0-BLH through
 R-CA0-BLN, R-CA0-BLS, R-CA0-BLU-V, R-CA0-BNR, and
 R-CA0-BUE. The resulting pool represented 20% of the
 final driver population. c) a pool of about 15,000 clones
 from non-normalized libraries CS0, CT0, CU0, CK0, and CX0
 and normalizing libraries CS0, CT0, CU0, CK0, and CX0,
 corresponding to plates R-CS0-CBD through R-CS0-CBO,
 R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CCM, R-CX0s-CCN
 through R-CX0s-CCX, R-CS0-CCB through R-CS0-CCW, R-CT0-BVN,
 R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVY,
 R-CU0-BUQ through R-CU0-BVL, R-CW0-BVY through R-CW0-BWP,
 R-CW0-BXN through R-CW0-BXO, R-CX0-BMO through R-CX0-BXM.
 The resulting pool represented 5% of the final driver
 population. d) a pool of about 5,000 clones (1,000 from
 non-normalized eye library CV0 and 4,000 from normalized
 eye library CV1) corresponding to plates R-CV0-BRH through
 R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through
 R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool
 represented about 5% of the final driver population. e) A
 pool of about 10,000 clones from subtracted library BS2,
 BV0 and BV0p (7-9.5 kb cDNA library fraction from rat
 whole embryo), and BX0 (0.5-7kb cDNA library fraction from
 rat whole embryo) corresponding to plates R-BS2-BDB
 through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0p-AOI
 through R-BV0p-AOX, and R-BX0-AQY through R-BX0-ASH. The
 resulting pool represented 5% of the final driver
 population. f) a pool of about 7,000 clones from the
 seven non-normalized libraries that make up the tracer
 including CV0, CZ0, DA0, DB0, DC0, DD0, and DE0
 corresponding to plates R-CY0-BXP through R-CY0-BXZ,
 R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ
 through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ
 through R-DB0-BZA, R-DC0-BZI through R-DC0-BZQ, R-DC0-CAY
 through R-DC0-CBA, R-DD0-BZR through R-DD0-CAJ,
 R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The
 resulting pool represented about 10% of the final driver
 population. g) a pool of about 2,000 clones from the pool
 of normalized libraries, CN0, that makes up the tracer.
 The corresponding plates are R-CN0-BKW through R-CN0-BLD,
 R-CN0-BLG, R-CN0-BLP through R-CN0-BLR, R-CN0-BLT,
 R-CN0-BLW-X, R-CN0-BMB, and R-CN0-BMF through R-CN0-BML.

This pool represented 5% of the final driver population.
 h) a pool of the 28 most abundant clones in the CN0 pool
 corresponding to the following addresses: bkx-a-09-0-UI,
 bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI,
 bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI,
 bkx-a-05-0-UI, bkx-a-06-0-UI, bkx-a-11-0-UI, bkx-c-06-0-UI,
 bkx-c-09-0-UI, bkx-d-10-0-UI, bla-a-01-0-UI,
 bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI,
 blb-a-12-0-UI, blb-f-02-0-UI, bic-a-11-0-UI,
 bic-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, bliq-h-04-0-UI,
 blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5%
 of the final driver population. 1) One abundant CN0 clone
 (corresponding to the address bkx-a-11-0-UI) was digested
 with Not I and Eco RI and the resulting insert was gel
 purified. This purified insert was added directly to the
 driver so that it represented 5% of the final driver
 population.

BASE COUNT 111 a 118 c 156 g 143 t
 ORIGIN
 TAG-LIB=UI-R-CNI
 TAG-TISSUE=brown adipose
 TAG-SEQ=TTGTC"

Query Match 2.5%; Score 19; DB 10; Length 528;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 CTTATCCACTCCCTGGAT 685
 |||||

DB 423 CTTATCCACTCCCTGGAT 405

RESULT 32

LOCUS BG383314 536 bp mRNA linear EST 12-MAR-2001
 DEFINITION 301304 MAMC LPIG Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BG383314

VERSION BG383314.1 GI:13307777

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 536)
 Authors: Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,
 and Keele, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

JOURNAL Unpublished (2000)

COMMENT Contact: Smith NPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt_rimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCGACGACGACG

Plate: 86 row: M column: 13

Seq primer: ATTTAGTGACACTATAG.

Location/Qualifiers

1..536

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC LPIG"

/tissue_type="pooled"

/lab_host="PH108"

/note="Vector: PCMV SPORt6; Site_1: XbaI; Site_2: XhoI;

FEATURES
 source

Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

BASE COUNT 93 a 173 c 168 g 102 t

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 536;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 442 AGCCTCTTGGACCTCTG 460

|||||

DB 95 AGCCTCTTGGACCTCTG 113

RESULT 33

LOCUS BF279171/c

DEFINITION BF279171 540 bp mRNA linear EST 07-MAR-2001
 GA_EB0037105f Gossypium arboreum 7-10 dpa fiber library Gossypium
 arboreum cDNA clone GA_EB0037105f, mRNA sequence.

ACCESSION BF279171

VERSION BF279171.2 GI:13248921

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.

REFERENCE 1 (bases 1 to 540)

Authors: Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
 D., Wood, T.C., Leslie, A., and Wilkins, T.A.

TITLE An integrated analysis of the genetics, development, and evolution
 of the cotton fiber

JOURNAL Unpublished (2000)

COMMENT On Nov 17, 2000 this sequence version replaced gi:11210241.

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATACGACTACTATAGG

High quality sequence start: 4

High quality sequence stop: 538.

FEATURES
 source

Location/Qualifiers

1..540

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db_xref="taxon:29729"

/clone="GA_EB0037105f"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue_type="Fibers isolated from bolls harvested 7-10

dpd"

/lab_host="E. coli"

/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 101 a 104 c 119 g 215 t

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 540;

Best Local Similarity 100.0%; Pred. No. 77;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 AAAATCAGCAACTCCCTCA 326

|||||

DB 309 AAAATCAGCAACTCCCTCA 291

RESULT 34

LOCUS AT731923/c

AT731923 563 bp mRNA linear EST 11-JUN-1999

```

DEFINITION  BNLIH11390 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
              to (297336) hypothetical protein [Arabidopsis thaliana], mRNA
              sequence.
ACCESSION   AI731923
VERSION     AI731923.1 GI:5050775
KEYWORDS    EST.
SOURCE      upland cotton.
ORGANISM    Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1 (bases 1 to 563)
AUTHORS    Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
TITLE      ESTs from developing cotton fiber
JOURNAL    Unpublished (1999)
COMMENT    Contact: Ben Burr
            Biology Department
            Brookhaven National Laboratory
            Upton, NY 11973, USA
            Tel: 516-344-3396
            Fax: 516-344-3407
            Email: burrb@bnl.bnl.gov
            Seq primer: T3 Primer.
FEATURES    Location/Qualifiers
            1..563
            /organism="Gossypium hirsutum"
            /db_xref="taxon:3635"
            /clone_lib="Six-day Cotton fiber"
            /tissue_type="Immature fiber"
            /dev_stage="Six days post anthesis"
            /lab_host="XLI-Blue"
            /note="Vector: pBluescript II KS+"
BASE COUNT  113 a 103 c 125 g 222 t
ORIGIN
            1..563
            /organism="Gossypium hirsutum"
            /db_xref="taxon:3635"
            /clone_lib="Six-day Cotton fiber"
            /tissue_type="Immature fiber"
            /dev_stage="Six days post anthesis"
            /lab_host="XLI-Blue"
            /note="Vector: pBluescript II KS+"

Query Match      2.5%; Score 19; DB 9; Length 563;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 AATATCAGCAACTCCCTCA 326
      |||||||
DB 262 AATATCAGCAACTCCCTCA 244

RESULT 35
AW141069
LOCUS    EST291092 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
DEFINITION  RGI8B01 5' and similar to DNA-directed RNA polymerase, mRNA
              sequence.
ACCESSION   AW141069
VERSION     AW141069.1 GI:6160837
KEYWORDS    EST.
SOURCE      Rattus sp.
ORGANISM    Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 569)
AUTHORS    Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    Other_ESTs: TC86667
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org

```

For clone availability, additional sequence and expression information related to this EST please check the TIGR Rat Gene Index (<http://www.tigr.org/tdb/rqi/rqi.html>). To order a clone contact the ATCC (<http://www.atcc.org/atcc.html>).

Seq primer: M13 Reverse.

FEATURES

source

1..569
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RG18B01"
 /clone_lib="Normalized rat kidney, Bento Soares"
 /note="Organ: Kidney; Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 134 a 152 c 169 g 112 t 2 others

ORIGIN

Query Match 2.5%; Score 19; DB 9; Length 569;

Best Local Similarity 100.0%; Pred. No. 78;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 CTTTCATCCACTCCCTGGAT 685

|||||

DB 102 CTTTCATCCACTCCCTGGAT 120

RESULT 36

AQ453959/c

LOCUS

DEFINITION HS_5165_B1_F02_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=741 Col=3 Row=L, DNA sequence.

ACCESSION AQ453959

VERSION AQ453959.1 GI:4595129

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 601)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering.bac.htm>) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 741 row: L column: 3

Seq primer: T7

Class: BAC ends

High quality sequence stop: 601.

Location/Qualifiers

1..601

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=741 Col=3 Row=L"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"

/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and

ECORI MethyIase. Size selected DNA was cloned into the pBACE3.6 vector at EORI sites" 13 others

BASE COUNT 126 a 162 c 136 g 164 t

ORIGIN

Query Match 2.5%; Score 19; DB 12; Length 601;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 524 GCCAGTACCAACCAACGCA 542

|||||

Db 147 GCCAGTACCAACCAACGCA 129

RESULT 37

AZ446530/c

LOCUS AZ446530 610 bp DNA linear GSS 04-OCT-2000

DEFINITION J06243D02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0243D02 F, DNA sequence.

ACCESSION AZ446530

VERSION AZ446530.1 GI:10597427

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 610)

REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0243 row: D column: 02

Seq primer: CGTTGTAACACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 610.

Location/Qualifiers

1..610

FEATURES

source

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0243D02"

/clone.lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 181 a 130 c 154 g 145 t

ORIGIN

Query Match 2.5%; Score 19; DB 12; Length 610;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 487 CCAGACACCAACCAACCTGTT 505

|||||

Db 314 CCAGACACCAACCAACCTGTT 296

RESULT 38

AZ858721/c

LOCUS AZ858721 683 bp DNA linear GSS 21-FEB-2001

DEFINITION 2M016AG05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC2M016AG05 F, DNA sequence.

ACCESSION AZ858721

VERSION AZ858721.1 GI:13052119

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 683)

REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0164 row: G column: 05

Seq primer: CGTTGTAACACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 683.

Location/Qualifiers

1..683

FEATURES

source

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M016AG05"

/clone.lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT
ORIGIN

195 a 162 c 160 g 165 t

Query Match 2.5%; Score 19; DB 12; Length 683;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GGAACCTGTTCTCCGCTG 33

|||||

Db 82 GGAACCTGTTCTCCGCTG 64

RESULT 39
B66149

LOCUS B66149 694 bp DNA linear GSS 21-JUN-1998
DEFINITION CIT-HSP-2024C22.TR CIT-HSP Homo sapiens genomic clone 2024C22, DNA sequence.

ACCESSION B66149

VERSION H66149.1 GI:2640127

KEYWORDS GSS

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 694)

AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden

,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.

and Venter,J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map

Building

Unpublished (1997)

Other_GSSs: CIT-HSP-2024C22.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES
source

Location/Qualifiers

1..694

/organism="Homo sapiens"

/db_xref="GDB:7046472"

/db_xref="taxon:9606"

/clone="2024C22"

/clone_lib="CIT-HSP"

/sex="Male"

/cell_type="Sperm"

/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:

HindIII"

BASE COUNT 148 a 137 c 121 g 288 t

ORIGIN

Query Match 2.5%; Score 19; DB 12; Length 694;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 ACATTTTGGTTGAAAT 312

|||||

Db 150 ACATTTTGGTTGAAAT 168

RESULT 40

AZ978692/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..739

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUCG2M0255E11"

/clone_lib="Mouse 10kb plasmid UUCG2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: PWD42hv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (gil4732114[gb]AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent *E. coli* XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 192 a 173 c 217 g 157 t

ORIGIN

Query Match 2.5%; Score 19; DB 12; Length 739;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 TGTATCTCTCCCTGCTGAC 640

|||||

Db 404 TGTATCTCTCCCTGCTGAC 386

AZ978692 739 bp DNA linear GSS 27-APR-2001
2M0255E11F Mouse 10kb plasmid UUCG2M library Mus musculus genomic
clone UUCG2M0255E11 F, DNA sequence.

ACCESSION AZ978692

VERSION AZ978692.1 GI:13849919

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 739)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0255 row: E column: 11

Seq primer: CGTTGTAACACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 739.

FEATURES

Location/Qualifiers

1..739

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUCG2M0255E11"

/clone_lib="Mouse 10kb plasmid UUCG2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: PWD42hv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (gil4732114[gb]AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent *E. coli* XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

```

RESULT 41
BF275075/c
LOCUS
DEFINITION
  BF275075 781 bp mRNA linear EST 07-MAR-2001
  GA_EB0023B03f Gossypium arboreum 7-10 dpa fiber library Gossypium
  arboreum cDNA clone GA_EB0023B03f, mRNA sequence.
ACCESSION
  BF275075
VERSION
  BF275075.2 GI:13248287
KEYWORDS
  EST.
SOURCE
  Gossypium arboreum.
  Gossypium arboreum.
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE
  1 (bases 1 to 781)
  Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
  D., Wood, T.C., Leslie, A. and Wilkins, T.A.
  An integrated analysis of the genetics, development, and evolution
  of the cotton fiber
  Unpublished (2000)
JOURNAL
  On Nov 17, 2000 this sequence version replaced gi:11206145.
COMMENT
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7286
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Seq primer: TAATACGACTCACTATAGGG
  High quality sequence stop: 698.
  Location/Qualifiers
  1..781
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B03f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 169 a 156 c 161 g 295 t
ORIGIN
  1..781
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B03f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

Query Match 2.5%; Score 19; DB 10; Length 781;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 AAATCAGCAACTCCCTCA 326
|||||
DB 424 AAATCAGCAACTCCCTCA 406

RESULT 42
BF275073/c
LOCUS
DEFINITION
  BF275073 782 bp mRNA linear EST 07-MAR-2001
  GA_EB0023B01f Gossypium arboreum 7-10 dpa fiber library Gossypium
  arboreum cDNA clone GA_EB0023B01f, mRNA sequence.
ACCESSION
  BF275073
VERSION
  BF275073.2 GI:13248285
KEYWORDS
  EST.
SOURCE
  Gossypium arboreum.
  Gossypium arboreum.
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE
  1 (bases 1 to 782)
  Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
  D., Wood, T.C., Leslie, A. and Wilkins, T.A.
  An integrated analysis of the genetics, development, and evolution
  of the cotton fiber
  Unpublished (2000)
JOURNAL
  On Nov 17, 2000 this sequence version replaced gi:11206145.
COMMENT
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7286
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Seq primer: TAATACGACTCACTATAGGG
  High quality sequence stop: 698.
  Location/Qualifiers
  1..781
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B01f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 169 a 156 c 161 g 295 t
ORIGIN
  1..781
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B01f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

Query Match 2.5%; Score 19; DB 10; Length 781;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 AAATCAGCAACTCCCTCA 326
|||||
DB 424 AAATCAGCAACTCCCTCA 406

```

```

JOURNAL
  Unpublished (2000)
COMMENT
  On Nov 17, 2000 this sequence version replaced gi:11206143.
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Seq primer: TAATACGACTCACTATAGGG
  High quality sequence stop: 761.
  Location/Qualifiers
  1..782
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B01f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 168 a 161 c 155 g 298 t
ORIGIN
  1..782
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B01f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

Query Match 2.5%; Score 19; DB 10; Length 782;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 AAATCAGCAACTCCCTCA 326
|||||
DB 424 AAATCAGCAACTCCCTCA 406

RESULT 43
BI829148/c
LOCUS
DEFINITION
  BI829148 796 bp mRNA linear EST 04-OCT-2001
  603079250F1 NTH_MGC_119 Homo sapiens cDNA clone IMAGE:5170885 5',
  mRNA sequence.
ACCESSION
  BI829148
VERSION
  BI829148.1 GI:15940698
KEYWORDS
  EST.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 796)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: Life Technologies, Inc.
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM1425 row: c column: 14
  High quality sequence stop: 796.
  Location/Qualifiers
  1..796
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:5170885"
  /clone_lib="NIH_MGC_119"
  /tissue_type="medulla"
  /lab_host="DH10B"
  /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
  Site_2: EcoRV (destroyed); RNA source normal medulla from

JOURNAL
  Unpublished (2000)
COMMENT
  On Nov 17, 2000 this sequence version replaced gi:11206143.
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Seq primer: TAATACGACTCACTATAGGG
  High quality sequence stop: 761.
  Location/Qualifiers
  1..782
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B01f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 168 a 161 c 155 g 298 t
ORIGIN
  1..782
  /organism="Gossypium arboreum"
  /strain="AKA"
  /cultivar="8400"
  /db_xref="taxon:29729"
  /clone_lib="GA_EB0023B01f"
  /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
  /tissue_type="Fibers isolated from bolls harvested 7-10
  dpa"
  /lab_host="E. coli"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

Query Match 2.5%; Score 19; DB 10; Length 782;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 AAATCAGCAACTCCCTCA 326
|||||
DB 424 AAATCAGCAACTCCCTCA 406

RESULT 43
BI829148/c
LOCUS
DEFINITION
  BI829148 796 bp mRNA linear EST 04-OCT-2001
  603079250F1 NTH_MGC_119 Homo sapiens cDNA clone IMAGE:5170885 5',
  mRNA sequence.
ACCESSION
  BI829148
VERSION
  BI829148.1 GI:15940698
KEYWORDS
  EST.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 796)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: Life Technologies, Inc.
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM1425 row: c column: 14
  High quality sequence stop: 796.
  Location/Qualifiers
  1..796
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:5170885"
  /clone_lib="NIH_MGC_119"
  /tissue_type="medulla"
  /lab_host="DH10B"
  /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
  Site_2: EcoRV (destroyed); RNA source normal medulla from

```

anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."

BASE COUNT 218 a 156 c 172 g 250 t
ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 796;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 AGTGCTTCAGAAATCCAGT 208

Db 182 AGTGCTTCAGAAATCCAGT 164

RESULT 44
BF275344/c

LOCUS 845 bp mRNA linear EST 07-MAR-2001
DEFINITION GA_Eb0023M18f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Eb0023M18f, mRNA sequence.

ACCESSION BF275344

VERSION BF275344.2 GI:13248335

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

1 (bases 1 to 845)

Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

, D., Wood, T.C., Leslie, A. and Wilkins, T.A.

An integrated analysis of the genetics, development, and evolution

of the cotton fiber

Unpublished (2000)

On Nov 17, 2000 this sequence version replaced gi:11206414.

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATACGACTGCTATACGG

High quality sequence stop: 798.

Location/Qualifiers

1..845

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db_xref="taxon:29729"

/clone="GA_EB0023M18f"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue_type="Fibers isolated from bolls harvested 7-10

dpa"

/lab_host="E. coli"

/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 183 a 172 c 175 g 313 t 2 others

ORIGIN

Query Match 2.5%; Score 19; DB 10; Length 845;

Best Local Similarity 100.0%; Pred. No. 88;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 AAATCAGCAACTCCCTCA 326

Db 424 AAATCAGCAACTCCCTCA 406

RESULT 45
BH363397/c

LOCUS 858 bp DNA linear GSS 03-DEC-2001

DEFINITION CH230-48F10.TVB CHORI-230 Segment 1 Rattus norvegicus genomic clone

CH230-48F10, DNA sequence.

ACCESSION BH363397

VERSION BH363397.1 GI:17294131

KEYWORDS GSS.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 858)

Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn

A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de

Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 EcoRI segment

Unpublished (1999)

Other_GSSs: CH230-48F10.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pjejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/orering_information.htm). BAC end

page: http://www.tigr.org/tdb/bac_ends/fat/bac_end_intro.html

Plate: 48 row: F column: 10

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..858

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-48F10"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by

Pieter de Jong"

BASE COUNT 310 a 161 c 143 g 244 t

ORIGIN

Query Match 2.5%; Score 19; DB 12; Length 858;

Best Local Similarity 100.0%; Pred. No. 88;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 613 GAAAGCACTGTATCTCTCC 631

Db 683 GAAAGCACTGTATCTCTCC 665

RESULT 46

BI462868/c

LOCUS 1095 bp mRNA linear EST 21-AUG-2001

DEFINITION 603207951F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5273461 5',

mRNA sequence.

ACCESSION BI462868

VERSION BI462868.1 GI:15253524

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 1095)
NIH-MGC http://www.ncbi.nlm.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapb@remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M11690 row: e column: 14
High quality sequence start: 3
High quality sequence stop: 628.
FEATURES
source
Location/Qualifiers
1..1095
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5273461"
/clone_lib="NIH_MGC_97"
/lab_host="DH103"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-gtgtttttttttttttt-3',
size-selected for average insert size 2.2 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT
310 a 245 c 349 g 191 t
ORIGIN

Query Match 2.5% Score 19: DB 10: Length 1095;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 19: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 327 TCCTGGATGTCACAAAGCTGA 345
|||||
DB 478 TCCTGGATGTCACAAAGCTGA 460

RESULT 47
AK013777
LOCUS
DEFINITION
Mus musculus adult male hippocampus cDNA, RIKEN full-length
non-receptor type 21, full insert sequence.
ACCESSION
AK013777
VERSION
1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain: C57BL/6J) adult male hippocampus cDNA to mRNA,
clone_lib: RIKEN full-length enriched mouse cDNA library
clone: 2900073M23.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
REFERENCE
2 (sites)
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes

```

```

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3 (sites)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hatada, A.,
Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, F., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 2343)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using triazole thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'-
GAGAGAGATTCGATTCGATTAATTAATTAATCCGCCGCC 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.
FEATURES
Location/Qualifiers
1..2343
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MGI:1905978"
/db_xref="taxon:10090"
/clone="2900073M23"
/sex="male"
/tissue_type="hippocampus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
360..713
/gene="Ptpn21"
360..713
/gene="Ptpn21"
gene
CDS

```

```

/note="data source:MGI", source key:MGI:134406,
evidence:ISS
protein tyrosine phosphatase, non-receptor type 21
putative"
/codon_start=1
/protein_id="BAB28990.1"
/db_xref="GI:12851270"
/translation="MPLPFGIKLRRRTVSSKSLVARIOLLNNEVEFTLSVEST
GQSLAVAGRLERKVTYFSIWNKQNRKVDLEPLKQLDKHALEPTVYRGVL
FYPSVSQLQDETR"
polyA_signal
2324..2329
/note="putative"
2343
/note="putative"
BASE COUNT 555 a 518 c 620 g 650 t
ORIGIN
polyA_site
2343
/note="putative"
Query Match 2.5% Score 19; DB 11; Length 2343;
Best Local Similarity 100.0%; Pred. NO. 1.2e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 CTCACCTGGGACCTCAGACT 59
|||||
Db 1229 CTCACCTGGGACCTCAGACT 1247

RESULT 48
AA980784/c 152 bp mRNA linear EST 27-MAY-1998
LOCUS IMAGE:1345710.5, mRNA sequence.
DEFINITION AA980784 Mus musculus cDNA clone
ACCESSION AA980784 GI:3159320
VERSION AA980784.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 152)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Sleptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,K.
The WashU-HMI Mouse EST Project
Unpublished (1996)
CONTACT: Marra M/Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through INM; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:698502
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 56.
Location/Qualifiers
1..152
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1345710"
/sex="male"
/tissue_type="mammary_gland_NbMMG"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site1: Not 1; Site2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)

```

```

primer [5',
TGTACCAATCTGAAGTGGGAGCGCGGGAATGGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 38 a 41 c 47 g 26 t
ORIGIN
Query Match 2.4% Score 18; DB 9; Length 152;
Best Local Similarity 100.0%; Pred. NO. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 TCCTGGATCTCAAGCTG 344
|||||
Db 54 TCCTGGATCTCAAGCTG 37

RESULT 49
BB055152 239 bp mRNA linear EST 25-JUN-2000
LOCUS BB055152 RIKEN full-length enriched 12 days embryo male wolffian
DEFINITION BB055152 RIKEN full-length enriched 12 days embryo male wolffian
musculus COP9 complex subunit 4 (COPS4) mRNA, mRNA sequence.
ACCESSION BB055152
VERSION BB055152.1 GI:8462300
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 239)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Koijima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Oho,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
CONTACT: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozaawa,Y., Muramatsu,M., Okazaki
Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

```

```

FEATURES
  source
    Location/Qualifiers
      1..239
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="6720477P07"
        /clone_lib="RIKEN full-length enriched, 12 days embryo
        male wolffian duct"
        /sex="male"
        /tissue_type="wolffian duct includes surrounding region"
        /dev_stage="12 days embryo"
        /lab_host="DH10B"
        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
        prepared and sequenced in Mouse Genome Encyclopedia
        Project of Genome Exploration Research Group in Riken
        Genomic Sciences Center and Genome Science Laboratory in
        RIKEN, Division of Experimental Animal Research in Riken
        contributed to prepare mouse tissues. 1st strand cDNA was
        primed with a primer [5'
        GACGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTTTNN 3']. cDNA was
        prepared by using trehalose thermo-activated reverse
        transcriptase and subsequently enriched for full-length by
        cap-trapper. Second strand cDNA was prepared with the
        primer adapter of sequence [5'
        GACGAGAGATTCGAGTTAATTAATCCCGCCCCCCC 3']. cDNA
        was cloned into the XhoI and BamHI sites. Vector: a
        modified pBluescript KS(+) after bulk excision from Lambda
        FLC 1. Cloning sites, 5' end: SalI; 3' end: BamHI"
      BASE COUNT      65 a  45 c  56 g
      ORIGIN
        1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 524 | 525 | 526 | 527 | 528 | 529 | 530 | 531 | 532 | 533 | 534 | 535 | 536 | 537 | 538 | 539 | 540 | 541 | 542 | 543 | 544 | 545 | 546 | 547 | 548 | 549 | 550 | 551 | 552 | 553 | 554 | 555 | 556 | 557 | 558 | 559 | 560 | 561 | 562 | 563 | 564 | 565 | 566 | 567 | 568 | 569 | 570 | 571 | 572 | 573 | 574 | 575 | 576 | 577 | 578 | 579 | 580 | 581 | 582 | 583 | 584 | 585 | 586 | 587 | 588 | 589 | 590 | 591 | 592 | 593 | 594 | 595 | 596 | 597 | 598 | 599 | 600 | 601 | 602 | 603 | 604 | 605 | 606 | 607 | 608 | 609 | 610 | 611 | 612 | 613 | 614 | 615 | 616 | 617 | 618 | 619 | 620 | 621 | 622 | 623 | 624 | 625 | 626 | 627 | 628 | 629 | 630 | 631 | 632 | 633 | 634 | 635 | 636 | 637 | 638 | 639 | 640 | 641 | 642 | 643 | 644 | 645 | 646 | 647 | 648 | 649 | 650 | 651 | 652 | 653 | 654 | 655 | 656 | 657 | 658 | 659 | 660 | 661 | 662 | 663 | 664 | 665 | 666 | 667 | 668 | 669 | 670 | 671 | 672 | 673 | 674 | 675 | 676 | 677 | 678 | 679 | 680 | 681 | 682 | 683 | 684 | 685 | 686 | 687 | 688 | 689 | 690 | 691 | 692 | 693 | 694 | 695 | 696 | 697 | 698 | 699 | 700 | 701 | 702 | 703 | 704 | 705 | 706 | 707 | 708 | 709 | 710 | 711 | 712 | 713 | 714 | 715 | 716 | 717 | 718 | 719 | 720 | 721 | 722 | 723 | 724 | 725 | 726 | 727 | 728 | 729 | 730 | 731 | 732 | 733 | 734 | 735 | 736 | 737 | 738 | 739 | 740 | 741 | 742 | 743 | 744 | 745 | 746 | 747 | 748 | 749 | 750 | 751 | 752 | 753 | 754 | 755 | 756 | 757 | 758 | 759 | 760 | 761 | 762 | 763 | 764 | 765 | 766 | 767 | 768 | 769 | 770 | 771 | 772 | 773 | 774 | 775 | 776 | 777 | 778 | 779 | 780 | 781 | 782 | 783 | 784 | 785 | 786 | 787 | 788 | 789 | 790 | 791 | 792 | 793 | 794 | 795 | 796 | 797 | 798 | 799 | 800 | 801 | 802 | 803 | 804 | 805 | 806 | 807 | 808 | 809 | 810 | 811 | 812 | 813 | 814 | 815 | 816 | 817 | 818 | 819 | 820 | 821 | 822 | 823 | 824 | 825 | 826 | 827 | 828 | 829 | 830 | 831 | 832 | 833 | 834 | 835 | 836 | 837 | 838 | 839 | 840 | 841 | 842 | 843 | 844 | 845 | 846 | 847 | 848 | 849 | 850 | 851 | 852 | 853 | 854 | 855 | 856 | 857 | 858 | 859 | 860 | 861 | 862 | 863 | 864 | 865 | 866 | 867 | 868 | 869 | 870 | 871 | 872 | 873 | 874 | 875 | 876 | 877 | 878 | 879 | 880 | 881 | 882 | 883 | 884 | 885 | 886 | 887 | 888 | 889 | 890 | 891 | 892 | 893 | 894 | 895 | 896 | 897 | 898 | 899 | 900 | 901 | 902 | 903 | 904 | 905 | 906 | 907 | 908 | 909 | 910 | 911 | 912 | 913 | 914 | 915 | 916 | 917 | 918 | 919 | 920 | 921 | 922 | 923 | 924 | 925 | 926 | 927 | 928 | 929 | 930 | 931 | 932 | 933 | 934 | 935 | 936 | 937 | 938 | 939 | 940 | 941 | 942 | 943 | 944 | 945 | 946 | 947 | 948 | 949 | 950 | 951 | 952 | 953 | 954 | 955 | 956 | 957 | 958 | 959 | 960 | 961 | 962 | 963 | 964 | 965 | 966 | 967 | 968 | 969 | 970 | 971 | 972 | 973 | 974 | 975 | 976 | 977 | 978 | 979 | 980 | 981 | 982 | 983 | 984 | 985 | 986 | 987 | 988 | 989 | 990 | 991 | 992 | 993 | 994 | 995 | 996 | 997 | 998 | 999 | 1000 | 1001 | 1002 | 1003 | 1004 | 1005 | 1006 | 1007 | 1008 | 1009 | 1010 | 1011 | 1012 | 1013 | 1014 | 1015 | 1016 | 1017 | 1018 | 1019 | 1020 | 1021 | 1022 | 1023 | 1024 | 1025 | 1026 | 1027 | 1028 | 1029 | 1030 | 1031 | 1032 | 1033 | 1034 | 1035 | 1036 | 1037 | 1038 | 1039 | 1040 | 1041 | 1042 | 1043 | 1044 | 1045 | 1046 | 1047 | 1048 | 1049 | 1050 | 1051 | 1052 | 1053 | 1054 | 1055 | 1056 | 1057 | 1058 | 1059 | 1060 | 1061 | 1062 | 1063 | 1064 | 1065 | 1066 | 1067 | 1068 | 1069 | 1070 | 1071 | 1072 | 1073 | 1074 | 1075 | 1076 | 1077 | 1078 | 1079 | 1080 | 1081 | 1082 | 1083 | 1084 | 1085 | 1086 | 1087 | 1088 | 1089 | 1090 | 1091 | 1092 | 1093 | 1094 | 1095 | 1096 | 1097 | 1098 | 1099 | 1100 | 1101 | 1102 | 1103 | 1104 | 1105 | 1106 | 1107 | 1108 | 1109 | 1110 | 1111 | 1112 | 1113 | 1114 | 1115 | 1116 | 1117 | 1118 | 1119 | 1120 | 1121 | 1122 | 1123 | 1124 | 1125 | 1126 | 1127 | 1128 | 1129 | 1130 | 1131 | 1132 | 1133 | 1134 | 1135 | 1136 | 1137 | 1138 | 1139 | 1140 | 1141 | 1142 | 1143 | 1144 | 1145 | 1146 | 1147 | 1148 | 1149 | 1150 | 1151 | 1152 | 1153 | 1154 | 1155 | 1156 | 1157 | 1158 | 1159 | 1160 | 1161 | 1162 | 1163 | 1164 | 1165 | 1166 | 1167 | 1168 | 1169 | 1170 | 1171 | 1172 | 1173 | 1174 | 1175 | 1176 | 1177 | 1178 | 1179 | 1180 | 1181 | 1182 | 1183 | 1184 | 1185 | 1186 | 1187 | 1188 | 1189 | 1190 | 1191 | 1192 | 1193 | 1194 | 1195 | 1196 | 1197 | 1198 | 1199 | 1200 | 1201 | 1202 | 1203 | 1204 | 1205 | 1206 | 1207 | 1208 | 1209 | 1210 | 1211 | 1212 | 1213 | 1214 | 1215 | 1216 | 1217 | 1218 | 1219 | 1220 | 1221 | 1222 | 1223 | 1224 | 1225 | 1226 | 1227 | 1228 | 1229 | 1230 | 1231 | 1232 | 1233 | 1234 | 1235 | 1236 | 1237 | 1238 | 1239 | 1240 | 1241 | 1242 | 1243 | 1244 | 1245 | 1246 | 1247 | 1248 | 1249 | 1250 | 1251 | 1252 | 1253 | 1254 | 1255 | 1256 | 1257 | 1258 | 1259 | 1260 | 1261 | 1262 | 1263 | 1264 | 1265 | 1266 | 1267 | 1268 | 1269 | 1270 | 1271 | 1272 | 1273 | 1274 | 1275 | 1276 | 1277 | 1278 | 1279 | 1280 | 1281 | 1282 | 1283 | 1284 | 1285 | 1286 | 1287 | 1288 | 1289 | 1290 | 1291 | 1292 | 1293 | 1294 | 1295 | 1296 | 1297 | 1298 | 1299 | 1300 | 1301 | 1302 | 1303 | 1304 | 1305 | 1306 | 1307 | 1308 | 1309 | 1310 | 1311 | 1312 | 1313 | 1314 | 1315 | 1316 | 1317 | 1318 | 1319 | 1320 | 1321 | 1322 | 1323 | 1324 | 1325 | 1326 | 1327 | 1328 | 1329 | 1330 | 1331 | 1332 | 1333 | 1334 | 1335 | 1336 | 1337 | 1338 | 1339 | 1340 | 1341 | 1342 | 1343 | 1344 | 1345 | 1346 | 1347 | 1348 | 1349 | 1350 | 1351 | 1352 | 1353 | 1354 | 1355 | 1356 | 1357 | 1358 | 1359 | 1360 | 1361 | 1362 | 1363 | 1364 | 1365 | 1366 | 1367 | 1368 | 1369 | 1370 | 1371 | 1372 | 1373 | 1374 | 1375 | 1376 | 1377 | 1378 | 1379 | 1380 | 1381 | 1382 | 1383 | 1384 | 1385 | 1386 | 1387 | 1388 | 1389 | 1390 | 1391 | 1392 | 1393 | 1394 | 1395 | 1396 | 1397 | 1398 | 1399 | 1400 | 1401 | 1402 | 1403 | 1404 | 1405 | 1406 | 1407 | 1408 | 1409 | 1410 | 1411 | 1412 | 1413 | 1414 | 1415 | 1416 | 1417 | 1418 | 1419 | 1420 | 1421 | 1422 | 1423 | 1424 | 1425 | 1426 | 1427 | 1428 | 1429 | 1430 | 1431 | 1432 | 1433 | 1434 | 1435 | 1436 | 1437 | 1438 | 1439 | 1440 | 1441 | 1442 | 1443 | 1444 | 1445 | 1446 | 1447 | 1448 | 1449 | 1450 | 1451 | 1452 | 1453 | 1454 | 1455 | 1456 | 1457 | 1458 | 1459 | 1460 | 1461 | 1462 | 1463 | 1464 | 1465 | 1466 | 1467 | 1468 | 1469 | 1470 | 1471 | 1472 | 1473 | 1474 | 1475 | 1476 | 1477 | 1478 | 1479 | 1480 | 1481 | 1482 | 1483 | 1484 | 1485 | 1486 | 1487 | 1488 | 1489 | 1490 | 1491 | 1492 | 1493 | 1494 | 1495 | 1496 | 1497 | 1498 | 1499 | 1500 | 1501 | 1502 | 1503 | 1504 | 1505 | 1506 | 1507 | 1508 | 1509 | 1510 | 1511 | 1512 | 1513 | 1514 | 1515 | 1516 | 1517 | 1518 | 1519 | 1520 | 1521 | 1522 | 1523 | 1524 | 1525 | 1526 | 1527 | 1528 | 1529 | 1530 | 1531 | 1532 | 1533 | 1534 | 1535 | 1536 | 1537 | 1538 | 1539 | 1540 | 1541 | 1542 | 1543 | 1544 | 1545 | 1546 | 1547 | 1548 | 1549 | 1550 | 1551 | 1552 | 1553 | 1554 | 1555 | 1556 | 1557 | 1558 | 1559 | 1560 | 1561 | 1562 | 1563 | 1564 | 1565 | 1566 | 1567 | 1568 | 1569 | 1570 | 1571 | 1572 | 1573 | 1574 | 1575 | 1576 | 1577 | 1578 | 1579 | 1580 | 1581 | 1582 | 1583 | 1584 | 1585 | 1586 | 1587 | 1588 | 1589 | 1590 | 1591 | 1592 | 1593 | 1594 | 1595 | 1596 | 1597 | 1598 | 1599 | 1600 | 1601 | 1602 | 1603 | 1604 | 1605 | 1606 | 1607 | 1608 | 1609 | 1610 | 1611 | 1612 | 1613 | 1614 | 1615 | 1616 | 1617 | 1618 | 1619 | 1620 | 1621 | 1622 | 1623 | 1624 | 1625 | 1626 | 1627 | 1628 | 1629 | 1630 | 1631 | 1632 | 1633 | 1634 | 1635 | 1636 | 1637 | 1638 | 1639 | 1640 | 1641 | 1642 | 1643 | 1644 | 1645 | 1646 | 1647 | 1648 | 1649 | 1650 | 1651 | 1652 | 1653 | 1654 | 1655 | 1656 | 1657 | 1658 | 1659 | 1660 | 1661 | 1662 | 1663 | 1664 | 1665 | 1666 | 1667 | 1668 | 1669 | 1670 | 1671 | 1672 | 1673 | 1674 | 1675 | 1676 | 1677 | 1678 | 1679 | 1680 | 1681 | 1682 | 1683 | 1684 | 1685 | 1686 | 1687 | 1688 | 1689 | 1690 | 1691 | 1692 | 1693 | 1694 | 1695 | 1696 | 1697 | 1698 | 1699 | 1700 | 1701 | 1702 | 1703 | 1704 | 1705 | 1706 | 1707 | 1708 | 1709 | 1710 | 1711 | 1712 | 1713 | 1714 | 1715 | 1716 | 1717 | 1718 | 1719 | 1720 | 1721 | 1722 | 1723 | 1724 | 1725 | 1726 | 1727 | 1728 | 1729 | 1730 | 1731 | 1732 | 1733 | 1734 | 1735 | 1736 | 1737 | 1738 | 1739 | 1740 | 1741 | 1742 | 1743 | 1744 | 1745 | 1746 | 1747 | 1748 | 1749 | 1750 | 1751 | 1752 | 1753 | 1754 | 1755 | 1756 | 1757 | 1758 | 1759 | 1760 | 1761 | 1762 | 1763 | 1764 | 1765 | 1766 | 1767 | 1768 | 1769 | 1770 | 1771 | 1772 | 1773 | 1774 | 1775 | 1776 | 1777 | 1778 | 1779 | 1780 | 1781 | 1782 | 1783 | 1784 | 1785 | 1786 | 1787 | 1788 | 1789 | 1790 | 1791 | 1792 | 1793 | 1794 | 1795 | 1796 | 1797 | 1798 | 1799 | 1800 | 1801 | 1802 | 1803 | 1804 | 1805 | 1806 | 1807 | 1808 | 1809 | 1810 | 1811 | 1812 | 1813 | 1814 | 1815 | 1816 | 1817 | 1818 | 1819 | 1820 | 1821 | 1822 | 1823 | 1824 | 1825 | 1826 | 1827 | 1828 | 1829 | 1830 | 1831 | 1832 | 1833 | 1834 | 1835 | 1836 | 1837 | 1838 | 1839 | 1840 | 1841 | 1842 | 1843 | 1844 | 1845 | 1846 | 1847 | 1848 | 1849 | 1850 | 1851 | 1852 | 1853 | 1854 | 1855 | 1856 | 1857 | 1858 | 1859 | 1860 | 1861 | 1862 | 1863 | 1864 | 1865 | 1866 | 1867 | 1868 | 1869 | 1870 | 1871 | 1872 | 1873 | 1874 | 1875 | 1876 | 1877 | 1878 | 1879 | 1880 | 1881 | 1882 | 1883 | 1884 | 1885 | 1886 | 1887 | 1888 | 1889 | 1890 | 1891 | 1892 | 1893 | 1894 | 1895 | 1896 | 1897 | 1898 | 1899 | 1900 | 1901 | 1902 | 1903 | 1904 | 1905 | 1906 | 1907 | 1908 | 1909 | 1910 | 1911 | 1912 | 1913 | 1914 | 1915 | 1916 | 1917 | 1918 | 1919 | 1920 | 1921 | 1922 | 1923 | 1924 | 1925 | 1926 | 1927 | 1928 | 1929 | 1930 | 1931 | 1932 | 1933 | 1934 | 1935 | 1936 | 1937 | 1938 | 1939 | 1940 | 1941 | 1942 | 1943 | 1944 | 1945 | 1946 | 1947 | 1948 | 1949 | 1950 | 1951 | 1952 | 1953 | 1954 | 1955 | 1956 | 1957 | 1958 | 1959 | 1960 | 1961 | 1962 | 1963 | 1964 | 1965 | 1966 | 1967 | 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | 2029 | 2030 | 2031 | 2032 | 2033 | 2034 | 2035 | 2036 | 2037 | 2038 | 2039 | 2040 | 2041 | 2042 | 2043 | 2044 | 2045 | 2046 | 2047 | 2048 | 2049 | 2050 | 2051 | 2052 | 2053 | 2054 | 2055 | 2056 | 2057 | 2058 | 2059 | 2060 | 2061 | 2062 | 2063 | 2064 | 2065 | 2066 | 2067 | 2068 | 2069 | 2070 | 2071 | 2072 | 2073 | 2074 | 2075 | 2076 | 2077 | 2078 | 2079 | 2080 | 2081 | 2082 | 2083 | 2084 | 2085 | 2086 | 2087 | 2088 | 2089 | 2090 | 2091 | 2092 | 2093 | 2094 | 2095 | 2096 | 2097 | 2098 | 2099 | 2100 | 2101 | 2102 | 2103 | 2104 | 2105 | 2106 | 2107 | 2108 | 2109 | 2110 | 2111 | 2112 | 2113 | 2114 | 2115 | 2116 | 2117 | 2118 | 2119 | 2120 | 2121 | 2122 | 2123 | 2124 | 2125 | 2126 | 2127 | 2128 | 2129 | 2130 | 2131 | 2132 | 2133 | 2134 | 2135 | 2136 | 2137 | 2138 | 2139 | 2140 | 2141 | 2142 | 2143 | 2144 | 2145 | 2146 | 2147 | 2148 | 2149 | 2150 | 2151 | 
```

```

LOCUS      W36972      296 bp      mRNA      linear      EST 11-SEP-1996
DEFINITION mb75c10.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:335250 5' similar to SW:RSPON_YEAST P13433 DNA-DIRECTED RNA
POLYMERASE MITOCHONDRIAL PRECURSOR ; mRNA sequence.
ACCESSION  W36972
VERSION     W36972.1 GI:1318358
KEYWORDS   EST
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 296)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
          Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
          Schellenberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
          Theising,B., Wylie,T., Lennon,C., Soares,B., Wilson,R. and
          Waterston,R.
TITLE      The WashU-HMMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
          WashU-HMMI Mouse EST Project
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: mouseest@watson.wustl.edu
          This clone is available royalty-free through LLNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          MGI:216650
          Seq primer: ETPRimer
          High quality sequence stop: 239.
          Location/Qualifiers
            1..296
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /clone="IMAGE:335250"
              /clone_lib="Soares mouse p3NMF19.5"
              /dev_stage="19.5 dpc total fetus"
              /lab_host="DH10B (ampicillin resistant)"
              /note="Vector: pT73D (Pharmacia) with a modified
              polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
              was primed with a Not I - oligo(dT) primer [5'
              TGTTACCAATCTCAAGTGGCGGCGCATTTTTTTTTTTT 3'],
              double-stranded cDNA was size selected, ligated to Eco RI
              adapters (Pharmacia), digested with Not I and cloned into
              the Not I and Eco RI sites of a modified pT73 vector
              (Pharmacia). Library went through one round of
              normalization to a Cot = 5. Library constructed by Bento
              Soares and M.Fatima Bonaudo. RNA was kindly provided by
              Dr. Minoru Ko (Wayne State University)."
```

```

BASE COUNT  72 a 104 c 65 g
ORIGIN

```

```

Query Match      2.4%; Score 18; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  667  CTTCACTCCACTCCCTGGA 684
      ||||||||||||||||
DB   136  CTTCACTCCACTCCCTGGA 153

```

```

RESULT  54
BF290469/c
LOCUS    BF290469      321 bp      mRNA      linear      EST 28-NOV-2000
DEFINITION EST455060 Rat Gene Index, normalized Rat. Rattus norvegicus cDNA
Rattus norvegicus cDNA clone RGIHV92 5' sequence, mRNA sequence.
ACCESSION BF290469
VERSION    BF290469.1 GI:11221539
KEYWORDS   EST
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
```

```

/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"
BASE COUNT  86 a 62 c 58 g 68 t 1 others
ORIGIN

Query Match      2.4%; Score 18; DB 9; Length 275;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  246  AGAAATGCTGAACATG 263
      ||||||||||||||||
DB   52  AGAAATGCTGAACATG 69

```

```

RESULT  52
AW312313
LOCUS    AW312313      276 bp      mRNA      linear      EST 08-JAN-2001
DEFINITION 4011 MARC lP1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  AW312313
VERSION     AW312313.1 GI:6728183
KEYWORDS   EST
SOURCE     pig.
ORGANISM   Sus scrofa
REFERENCE  1 (bases 1 to 276)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
          Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
          Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
          and Keele,J.W.
          Design and use of two pooled tissue normalized cDNA libraries for
          EST discovery in swine
          Unpublished (2000)
          Contact: Smith TPL
          USDA, ARS, US Meat Animal Research Center
          PO Box 166, Clay Center, NE 68333-0166, USA
          Tel: 402 762 4366
          Fax: 402 762 4390
          Email: smith@email.marc.usda.gov
          Single pass sequencing. Bases called and trimmed with phred
          v0.980904.e. Vector identified by cross_match with the -minscore 20
          and -mismatch 12 options.
          PCR Primers
          FORWARD: AGCAACAGCTATGACCAT
          BACKWARD: GTTTCCTCAGTCACGACG
          Plate: 136 row: B column: 8
          Seq primer: ATTAGGTGACACTATAG.
          Location/Qualifiers
            1..276
              /organism="Sus scrofa"
              /db_xref="taxon:9823"
              /clone_lib="MARC lP1G"
              /tissue_type="pooled"
              /lab_host="DH10B"
              /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
              Library made from pooled tissue from day 11, 13, 15, 20,
              and 30 embryos."
```

```

BASE COUNT  69 a 72 c 104 g 31 t
ORIGIN

```

```

Query Match      2.4%; Score 18; DB 9; Length 276;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  226  GCAGAGCCCGCAGGAGC 243
      ||||||||||||||||
DB   220  GCAGAGCCCGCAGGAGC 237

```

```

RESULT  53
W36972
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 321)
Malek, R.L., Cho, J., Lee, Y., Karamecheva, S., Parvizi, B., Pertea, G.,
Sultana, R., Tsai, J., White, J., Quackenbush, J., and Lee, N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Unpublished (2000)
Contact: Lee, N.H.
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.

Location/Qualifiers

1. 321

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone_lib="RGLHV92"

/clone_lib="Rat Gene Index, normalized rat, Rattus

norvegicus cDNA"

/tissue_type="mixed tissue"

/lab_host="DH5-alpha"

/note="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;

Generation of ROV, RBR, RKI, RLI, RPL, REM, RNU, RSP

Combination of RHE, RPE, RPN"

61 a 90 c 92 g 78 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 2.2e+02; Length 321;

Matches 18: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 613 GAAAGCACTGATCTCTC 630

Db 194 GAAAGCACTGATCTCTC 177

RESULT 55

AM662992/c

LOCUS

DEFINITION

hB3h02.y1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2969427 5'

similar to TR:075139 075139 KIAA0644 PROTEIN. i, mRNA sequence.

AM662992

AM662992.1 GI:7455533

EST.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 323)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Other ESTs: hB3h02.x1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/iresources.shtml

Seq primer: 40bp from Gibco

High quality sequence stop: 306.

Location/Qualifiers

1. 323

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:2969427"

/clone_lib="NCI_CGAP_GU1"

/tissue_type="2 pooled high-grade transitional cell

tumors"

/lab_host="DH108"

/note="Organ: genitourinary tract; Vector: pCMV-SPORT6;

Site_1: SalI; Site_2: NotI; Cloned unidirectionally.

Primer: Oligo dt. Library constructed by Life

Technologies."

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 2.2e+02; Length 323;

Matches 18: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 494 CACCAGCTGTGCGTC 511

Db 212 CACCAGCTGTGCGTC 195

RESULT 56

BB491192

LOCUS

DEFINITION

musculus cDNA clone D53001912 3', similar to AF071314 Mus musculus

COP9 complex subunit 4 (COPS4) mRNA, mRNA sequence.

BB491192

BB491192.1 GI:9449819

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 323)

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci

P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,

Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.

Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.

Oho, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata

Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sodabe, Y., Sugahara, Y.

Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya

T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamakata, I.

Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino

M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp

URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki

N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki

Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers
1. 323
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D530019N12"
/clone_lib="RIKEN full-length enriched, 13 days embryo stomach"
/tissue_type="stomach"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',
GAGAGAGATTCTCGAGTTAATTAATTCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified phluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 74 a 78 c 64 g 107 t
ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GCTTTGGAACCTGTTCT 27
|||||
Db 189 GCTTTGGAACCTGTTCT 206
|||||

RESULT 57
AA751893/c
LOCUS
DEFINITION
cDNA clone 96AS0752, mRNA sequence.
ACCESSION
AA751893.1 GI:2798599
VERSION
EST.
KEYWORDS
Oryza sativa.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 332)
Nahm,B.H., Kim,W.Y., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.F., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeunsun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. 332
/organism="Oryza sativa"

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers
1. 323
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D530019N12"
/clone_lib="RIKEN full-length enriched, 13 days embryo stomach"
/tissue_type="stomach"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',
GAGAGAGATTCTCGAGTTAATTAATTCGCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified phluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 74 a 78 c 64 g 107 t
ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 323;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GCTTTGGAACCTGTTCT 27
|||||
Db 189 GCTTTGGAACCTGTTCT 206
|||||

RESULT 57
AA751893/c
LOCUS
DEFINITION
cDNA clone 96AS0752, mRNA sequence.
ACCESSION
AA751893.1 GI:2798599
VERSION
EST.
KEYWORDS
Oryza sativa.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 332)
Nahm,B.H., Kim,W.Y., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.F., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeunsun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. 332
/organism="Oryza sativa"

/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="96AS0752"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
BASE COUNT 58 a 110 c 84 g 78 t
ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 332;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 727 AACCCAGCTGCAACCT 744
|||||
Db 327 AACCCAGCTGCAACCT 310
|||||

RESULT 58
AA671815
LOCUS
DEFINITION
v102a07.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:962964 5' similar to SW:RPOM_YEAST P13433 DNA-DIRECTED RNA POLYMERASE MITOCHONDRIAL PRECURSOR ;, mRNA sequence.
ACCESSION
AA671815 GI:2543894
VERSION
EST.
KEYWORDS
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 335)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:551756
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 292.
Location/Qualifiers
1. 335
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:962964"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T30-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(gt) primer [5',
TGTTACCAATCTGAAGTGGGAGCGCGCGGATGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 73 a 116 c 80 g 66 t

ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 335;

Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 667 CTTCACTCACTCCCTGGG 684

|||||

Db 221 CTTCACTCACTCCCTGGG 238

RESULT 59

BB224526

LOCUS

DEFINITION

musculus cDNA clone A530086P22 3' similar to NM_001706 Homo sapiens B-cell CLL/lymphoma 6 (zinc finger protein 51) (BCL6) mRNA, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 335)

REFERENCE

AUTHORS

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Okakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, H., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermostable enzymes for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1..335

TITLE

JOURNAL

MEDLINE

96026280

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="A530086P22"

/cclone.lib="RIKEN full-length enriched, adult male aorta and vein"

/sex="male"

/tissue_type="aorta and vein"

/dev_stage="adult"

/lab_host="DH10B"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',

GAGAGAGAGAGGATCCCAAGGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using triazole thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adaptor of sequence [5', GAGAGAGAGATTCGAGTTAATTAATTCCTCCCTCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda

FLC I."

BASE COUNT 88 a 76 c 62 g 109 t

ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 335;

Best Local Similarity 100.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 CTGCAGAGGAGATATGCT 652

|||||

Db 19 CTGCAGAGGAGATATGCT 36

RESULT 60

AA350438

LOCUS

DEFINITION

AA350438

VERSION

KEYWORDS

SOURCE

ORGANISM

Human

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 336)

REFERENCE

AUTHORS

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Whi, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodde, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, P.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hung, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

COMMENT

Other ESTs: EST57698 THC173165
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers
 1..336

/organism="Homo sapiens"
 /db_xref="ATCC (inhost):104035"
 /db_xref="taxon:9606"
 /clone_lib="infant brain"
 /sex="female"
 /dev_stage="infant"
 /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
 Site_2: NotI"

BASE COUNT
 ORIGIN

112 a 70 c 76 g 78 t

Query Match

Best Local Similarity 2.4%; Score 18; DB 9; Length 336;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 AGAATTGCTGACCAATG 263

|||||

Db 81 AGAATTGCTGACCAATG 98

RESULT 61

AA285185

LOCUS

DEFINITION AA285185 341 bp mRNA linear EST 15-DEC-1997
 IMAGE:701843 5', mRNA sequence.

ACCESSION AA285185

VERSION AA285185.1

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Hillier, L., Allen, M., Rowles, L., Dubuque, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, R.,
 Schellenger, K., Stepien, M., Tan, P., Theisling, B., White, Y., Wyllie,
 T., Waterston, R. and Wilson, R.
 WashU-Merck EST Project 1997

TITLE

Unpublished (1997)

Contact: Wilson RK

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available royalty-free through LML; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 308

FEATURES

Location/Qualifiers

1..341
 /organism="Homo sapiens"
 /db_xref="GDB:5742514"
 /db_xref="taxon:9606"
 /clone_lib="IMAG:701843"
 /clone_lib="Stratagene schizo brain S11"
 /sex="male"
 /tissue_type="schizophrenic brain S-11 frontal lobe"

/dev_stage="34 years old"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site_1: EcoRI; Library

constructed from S-11 frontal lobe, male, 34 years old,

50% caucasian, 50% Aleutian. Schizophrenic suicide.

Random primed into EcoRI site of 2AP II vector. Mass

excised. Custom library. Avg insert length 1.4kb.

Material obtained by Johnston N., Torrey, E.F., Volken R.,

and the Stanley Neuropathology Consortium - Analysis of

KNAs from the Brains of Individuals with psychiatric

Diseases (Unpublished) Stanley Neurovirology Laboratory,

Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 108 a 77 c 57 g 99 t

ORIGIN

Query Match

Best Local Similarity 2.4%; Score 18; DB 9; Length 341;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 AGAATTGCTGACCAATG 263

|||||

Db 322 AGAATTGCTGACCAATG 339

RESULT 62

AI372485

LOCUS

DEFINITION AI372485 343 bp mRNA linear EST 12-JAN-1999

FUHI899, mRNA sequence.

ACCESSION AI372485

VERSION AI372485.1

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 343)

AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult

, C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D., White

, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,

Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald

, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,

Ghosh, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,

Kelley, J.C., Liu, L.-F., Marmaros, S.M., Merrick, J.M.,

Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Shall, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,

Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,

, Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,

Kuhsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,

Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon

, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and

Venter, J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

Other ESTs: EST175319 EST175321 EST175322 THC290418

Contact: HGI (Human Gene Index)

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-0200

Fax: (301)-838-0208

Email: hgi@tigr.org

Location/Qualifiers

1..343
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="FUHI899"
 /clone_lib="infant brain, Bento Soares"
 /lab_host="E. coli DH5-alpha"

/note="vector: BA, M13-derived; Site_1: HindIII; Site_2: NotI; The infant brain library, constructed by Bento Soares, Columbia University, was oligo-(dT) primed and directionally cloned into an M13-derived plasmid using total brain mRNA from a 72-day old human female afflicted with spinal muscular atrophy."

BASE COUNT
ORIGIN

114 a 72 c 79 g 78 t

Query Match 2.4% Score 18; DB 9; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 AGAAATTGCTGACAAATG 263

|||||
Db 85 AGAAATTGCTGACAAATG 102

RESULT 63
A0918301/c
LOCUS
DEFINITION
A0918301
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

351 bp DNA linear GSS 21-DEC-1999
RPCI-23-271B8.TJ RPCI-23 Mus musculus genomic clone RPCI-23-271B8,
DNA sequence.
A0918301
A0918301.1 GI:6607303
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 351)
Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinret
B., Levis,M., McGann,S., Tsegaye,G., Geer,K., Kroll,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-271B8.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@jeng.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 271 row: B column: 8
Seq primer: SP6
Class: BAC ends.

FEATURES
source

1. .351 Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-271B8"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
71 a 102 c 107 g 70 t 1 others

BASE COUNT
ORIGIN

71 a 102 c 107 g 70 t

Query Match 2.4% Score 18; DB 12; Length 351;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 TCCTTGAGAACTGAAGG 179

|||||
Db 41 TCCTTGAGAACTGAAGG 24

RESULT 64

W36801/c

LOCUS

DEFINITION

W36801

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

W36801

mb90d01.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

IMAGE:336673 5', mRNA sequence.

W36801

W36801.1 GI:1318723

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 359)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:218073

Seq primer: mob.REGA-ET

High quality sequence stop: 340.

Location/Qualifiers

1. .359

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:336673"

/clone_lib="Soares mouse p3NMF19.5"

/dev_stage="19.5 dpc total fetus"

/lab_host="DH10B (ampicillin resistant)"

/note="Vector: pT73D (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5']

TCTTACCAATCTCACTGGAGCGCGCATTTTTTTTTTTTTTTT 3',

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo. RNA was kindly provided by

Dr. Minoru Ko (Wayne State University)."

76 a 89 c 112 g 82 t

BASE COUNT

ORIGIN

Query Match 2.4% Score 18; DB 10; Length 359;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 TCCTGGATGTCAAAGCTG 344

|||||

Db 298 TCCTGGATGTCAAAGCTG 281

```

RESULT 65
F06455
LOCUS
DEFINITION
HSC18G111 normalized infant brain cDNA Homo sapiens cDNA clone
c-18g11, mRNA sequence.
EST.
ACCESSION F06455.1 GI:672030
VERSION F06455
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 361)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Degivnes,
M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabakchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534
CONTACT: Genethon
Genethon-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genethon@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-18g11
Seq primer: (-21)M13 universal.
FEATURES
Location/Qualifiers
1..361
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-18g11"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/notes="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex:Female; dev_stage=3 months old;
Isolate:muscular atrophy patient; tissue_type=total brain
cloned 5' -> 3' into the HindIII -> NotI sites of the
Jafmid BA vector. Clone library from B.Soaress, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 116 a 75 c 64 g 103 t 3 others
ORIGIN
Query Match 2.4%; Score 18; DB 10; Length 361;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 AGAATTGCTGAACAATG 263
|||||
Db 293 AGAATTGCTGAACAATG 310

RESULT 66
W07561
LOCUS
DEFINITION
za98h03.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
IMAGE:300629 5', mRNA sequence.
W07561
ACCESSION W07561.1 GI:1281564
VERSION W07561
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Murinae;
Rattus.
1 (bases 1 to 370)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping

```

```

REFERENCE
AUTHORS Hillier,L., Clark,M., Dubuque,T., Elliston,K., Hawkins,M., Holman,
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Kifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,
R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.MEGA-ET
High quality sequence stop: 239.
FEATURES
Location/Qualifiers
1..369
/organism="Homo sapiens"
/db_xref="GDB:1245553"
/db_xref="taxon:9606"
/clone="IMAGE:300629"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: lung; Vector: p773D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTITTTT-3'] ,
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."
BASE COUNT 109 a 71 c 91 g 92 t 6 others
ORIGIN
Query Match 2.4%; Score 18; DB 10; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 AGAATTGCTGAACAATG 263
|||||
Db 9 AGAATTGCTGAACAATG 26

RESULT 67
BG379573
LOCUS
DEFINITION
UI-R-CSO-bti-e-09-0-UI-sl UI-R-CSO Rattus norvegicus cDNA clone
UI-R-CSO-bti-e-09-0-UI 3', mRNA sequence.
BG379573
ACCESSION BG379573
VERSION BG379573.1 GI:13304045
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 370)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping

```

University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoures@blue.weeg.uiowa.edu
 The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the normalized rat heart pool library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-Yes.

FEATURES

Source

Location/Qualifiers

1. 370
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-CS0-bti-e-09-0-01"
 /clone_lib="UI-R-CS0"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pMT3D-Pac (Pharmacia) with a modified polylinker. Site1: Not I; Site2: Eco RI; The UI-R-CS0 library is a normalized library constructed from the following rat heart tissues: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, embryonic day 200. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG-LIB=UI-R-CS0
 TAG-TISSUE=rat heart pool
 TAG_SEQ=ATAAGATAAC"

BASE COUNT 95 a 89 g 104 t

ORIGIN

Query Match 2.4%; Score 18; DB 10; Length 370;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 613 GAAAGCACTGTATCCTC 630

|||||

DB 325 GAAAGCACTGTATCCTC 342

RESULT 68

AA199532

LOCUS

DEFINITION

IMAGE:657488 5', mRNA sequence.

AA199532

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

TITLE

WASHU-HMI Mouse EST Project

Unpublished (1996)

Contact: Maria M/Mouse EST Project

WASHU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.

MGI:403336

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 349.

FEATURES

Source

Location/Qualifiers

1. 374
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:657488"
 /clone_lib="Soares mouse 3NME12 5"
 /sex="unknown"
 /tissue_type="fetus"
 /dev_stage="12.5dpc total fetus"
 /lab_host="DH10B"
 /note="Organ: whole fetus; Vector: pMT3D-Pac (Pharmacia) with a modified polylinker. Site1: Not I; Site2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGCAGCGCGCTTTATTTTTTTTTTTT 3'], on total mouse RNA (provided by Minoru Ko, Wayne State Univ.); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pMT3D vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 115 a 69 c 88 g 102 t

ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 374;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 CTCAGAGTCTCTCTTGA 69

|||||

DB 209 CTCAGAGTCTCTCTTGA 226

RESULT 69

BF429883/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

TITLE

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@meat.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

v0.980904 e. Vector identified by cross_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: GGAACACGCTATGACCATG

BACKWARD: GTTTCCTCAGTCACGAC

Seq primer: AATTAACCTCTCACTAAAGG.

FEATURES
source

Location/Qualifiers

1. .375
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC BSM"
 /tissue_type="Skeletal muscle"
 /lab_host="XL0LR"
 /note="Vector: Uni-ZAP XR; Site_1: EcoRI; Site_2: XhoI;
 Library obtained from Stratagene, catalog #937721. Library
 made from skeletal muscle of a two year old Holstein cow."
 64 a 101 c 93 g 117 t

BASE COUNT
 ORIGIN

Query Match 2.4%; Score 18; DB 10; Length 375;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 CAGCCAAATCATCAACAA 585

Db 36 CAGCCAAATCATCAACAA 19

RESULT 70

AI005438

LOCUS
 DEFINITION
 IMAGE:1625667 3', mRNA sequence.

ACCESSION

AI005438

VERSION

AI005438.1

KEYWORDS

EST.

SOURCE

ORGANISM

Human.

REFERENCE

1 (bases 1 to 376)

AUTHORS

NCI-CGAP

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through (LNL) ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 601 Std Error: 0.00

Seq primer: primer name ambiguous

High quality sequence stop: 352.

Location/Qualifiers

1. .376

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI-CGAP GCBI) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo. "

84 a 97 c 107 g 88 t

BASE COUNT

ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 376;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 GGCCCATCATTCGCCAGA 428

Db 314 GGCCCATCATTCGCCAGA 331

RESULT 71

T33416

LOCUS

DEFINITION

EST57781 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA

sequence.

T33416

ACCESSION

T33416.1

VERSION

GI:615514

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 380)

AUTHORS

Adams, M.D., Kerlavage, A.K., Fleischmann, R.D., Fuldner, R.A., Bult

, C.J., Lee, N., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White

, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.W., Clayton, R.A.,

Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., FitzGerald

, L.M., FitzHugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodek, A.,

Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkley, P.S., Kelley, J.M.,

Kline, K.M., Kelley, J.C., Liu, L.-T., Marmaros, S.M., Merrick, J.M.,

Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,

Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,

Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.-J.,

Dimke, D., Feng, P., Ferrrie, A., Fischer, C., Hastings, G.A., He, W.-W.,

Hu, J.-S., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L.,

Kunsch, C., Ji, H., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei

, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon

, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and

Venter, J.C.

Initial Assessment of Human Gene Diversity and Expression Patterns

Based Upon 83 Million Basepairs of cDNA Sequence

Nature 377, 3-174 (1995)

Other ESTs: EST57780

Contact: Venter, JC

The Institute for Genomic Research

932 Clopper Rd, Gaithersburg, MD 20878

Tel: 3018659056

Fax: 3018699423

Email: tdbinfo@tdb.tigr.org

For clone availability, additional sequence and expression

information related to this EST, please contact the TIGR Database

(tdbinfo@tdb.tigr.org)

Seq primer: M13 Reverse.

Location/Qualifiers

1. .380

/organism="Homo sapiens"

/db_xref="ATCC (inhost):104057"

/db_xref="taxon:9606"

/clone_lib="Human Brain"

/note="Organ: brain"

BASE COUNT 112 a 75 c 95 g 98 t

ORIGIN

Query Match 2.4%; Score 18; DB 10; Length 380;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 AGAAATTCGTGAACAATG 263

Db 17 AGAAATTCGTGAACAATG 34

RESULT 72

AA856299

REFERENCE

1 (bases 1 to 390)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 2023

TITLE
JOURNAL
COMMENT

High quality sequence stops: 342 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 2023 Std Error: 0.00

Seq primer: M13RP1
 High quality sequence stop: 342.

FEATURES

source
 1..390
 /organism="Homo sapiens"
 /db_xref="GDB:396313"
 /db_xref="taxon:9606"
 /clone="IMAGE:23966"
 /clone_lib="Soares infant brain INIB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: Lfamid BA; Site_1: Not
 I; Site_2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5,
 ACTGGAAATTCGCGCCGACGAATTTTCTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the Lfamid BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
 ORIGIN

124 a 83 c 80 g 102 t 1 others

Query Match 2.4%; Score 18; DB 10; Length 390;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 AGAATTCTGACAAATG 263

|||||

Db 164 AGAATTCTGACAAATG 181

RESULT 75

Al164059/c

LOCUS

DEFINITION At164059 Hybrid aspen plasmid library Populus tremula x Populus

tremuloides cDNA 5', mRNA sequence.

ACCESSION

Al164059

VERSION

Al164059.1

KEYWORDS

SOURCE

ORGANISM

Populus tremula x Populus tremuloides.
 Populus tremula x Populus tremuloides.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.

1 (bases 1 to 392)
 Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg
 A., Amini, B., Bhalarao, R., Larsson, M., Villarroel, R., Van Montagu
 M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P.,
 Uhlen, M., Sundberg, B. and Lundberg, J.

Gene discovery in the wood-forming tissues of poplar: Analysis of
 5,692 expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)

MEDLINE
COMMENT

99007314
 Contact: Sterky F
 Department of Biotechnology
 Royal Institute of Technology (KTH)
 Teknikringen 34, S-100 44 STOCKHOLM, Sweden
 Tel: +46 8 790 8287
 Fax: +46 8 24 54 52
 Email: fredrik@biochem.kth.se

PCR Primers
 FORWARD: AAAGGGGATGCTCTGCAAGGCG
 BACKWARD: GCTTCGGCTCTGATGTTGTG
 Seq primer: CGTTGAAACGACGCCAG
 High quality sequence stop: 392.

FEATURES

source
 1..392
 /organism="Populus tremula x Populus tremuloides"
 /db_xref="taxon:47664"
 /clone_lib="Hybrid aspen plasmid library"
 /tissue_type="Cambial region"
 /dev_stage="1.5 m actively growing tree"
 /lab_host="E.coli"

/note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
 Cambial region tissues, including developing xylem, the
 meristematic cambial zone and the developing and mature
 phloem, was harvested from 1.5 m actively growing trees.
 cDNA was prepared and cloned into lambda gt2a. DNA was
 isolated and subcloned into pBluescript SK using SalI and
 NotI restriction enzymes."

BASE COUNT 130 a 70 c 99 g 92 t 1 others

ORIGIN

Query Match 2.4%; Score 18; DB 9; Length 392;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 TCCTCTCTTGACAAATCT 76

|||||

Db 67 TCCTCTCTTGACAAATCT 50

Search completed: August 6, 2002, 19:49:50
 Job time: 5054 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 17:28:26 : Search time 212.65 Seconds
(without alignments)
6023.128 Million cell updates/sec

Title: US-10-020-139-1_COPY_48_793

Perfect score: 746

Sequence: 1 CATGCTTCAGCTTTGGAAC.....AACCCAGCTGCAAAACCTCA 746

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 250 summaries

Database : N_Geneseq_032802.*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	746	100.0	824	19 AAV37699	Human parotid secr
2	746	100.0	1028	19 AAV44759	Human parotid secr
3	746	100.0	1049	21 AAC58379	Human PRO1025 nucl
4	746	100.0	1058	21 AAB87727	Human secreted pro
5	746	100.0	1058	22 AAF64009	cDNA encoding huma
6	404	54.2	1041	22 AAH96866	Human protein enco
7	218	29.2	538	19 AAV44761	Expressed sequence
8	211	28.3	406	19 AAV44764	Expressed sequence
9	182	24.4	493	19 AAV44765	Expressed sequence

10	177	23.7	317	21 AAC23485	Human secreted pro
11	159	21.3	449	19 AAV44760	Expressed sequence
12	159	21.3	7524	22 AK89798	Human digestive sy
13	154	20.6	581	22 ABA60313	Human foetal liver
14	154	20.6	581	22 AAK08593	Human brain expres
15	154	20.6	581	22 AAK34475	Human bone marrow
16	154	20.6	581	22 AA140197	Probe #8883 used t
17	153	20.5	153	22 ABA72852	Human foetal liver
18	153	20.5	153	22 AAK21285	Human brain expres
19	153	20.5	153	22 AAK47441	Human bone marrow
20	153	20.5	153	22 AA153276	Probe #21962 used
21	150	20.1	374	19 AAV44763	Expressed sequence
22	110	14.7	359	19 AAV44762	Expressed sequence
23	28	3.8	39	19 AAV44769	PCR primer for hum
24	25	3.4	43	19 AAV44771	PCR primer for hum
25	24	3.2	395	19 AAV44766	Expressed sequence
26	22	2.9	36	19 AAV44770	PCR primer for hum
27	21	2.8	1730	22 AAH18450	Human cDNA sequenc
28	20	2.7	360	19 AAV44768	Expressed sequence
29	20	2.7	4692	19 AAV04633	Porcine phospholip
30	20	2.7	4692	20 AAV74099	Porcine G-protein
31	20	2.7	4692	21 AA286812	Pig p101 regulator
32	19	2.5	425	22 ABA11053	Human nervous syst
33	19	2.5	451	22 ABA14447	Human nervous syst
34	19	2.5	766	22 ABA14448	Human nervous syst
35	19	2.5	2957	19 AAV18462	Human granulocytic
36	19	2.5	3636	23 AAS91657	DNA encoding novel
37	19	2.5	59060	22 AAK85994	Human immune/haema
38	18	2.4	846	23 ABL17739	Drosophila melanog
39	18	2.4	1173	23 AAS53405	Haemophilus influe
40	18	2.4	1274	22 AAB73917	Human PD2 protein
41	18	2.4	2265	22 AAS46228	Human DNA encoding
42	18	2.4	2345	20 AAX18951	Human DNA encoding
43	18	2.4	2643	23 ABL17740	Human PI3K-1 encod
44	18	2.4	3224	18 AAT91872	Drosophila melanog
45	18	2.4	3839	23 ABL09202	Rat Yf521 gene. R
46	18	2.4	4331	23 ABL18573	Drosophila melanog
47	18	2.4	4360	23 ABL12151	Drosophila melanog
48	18	2.4	4372	22 ABL16649	Human cDNA sequenc
49	18	2.4	5098	23 ABL17738	Drosophila melanog
50	18	2.4	6642	23 ABL08533	Drosophila melanog
51	18	2.4	6846	23 ABL12150	Drosophila melanog
52	18	2.4	6872	23 ABL18572	Drosophila melanog
53	18	2.4	8115	22 AAS32860	Drosophila melanog
54	18	2.4	8116	22 AAS32861	Human genomic DNA
55	18	2.4	8844	23 ABL08532	Human genomic DNA
56	18	2.4	92407	22 AAF28549	Drosophila melanog
57	17	2.3	249	22 ABA51549	Genomic fragment #
58	17	2.3	249	22 ABA36520	Human breast cell
59	17	2.3	350	21 AAS56190	Probe #14986 for g
60	17	2.3	395	22 AA180514	Eucalyptus grandis
61	17	2.3	454	22 ABA46449	Human polynucleoti
62	17	2.3	454	22 ABA26652	Human breast cell
63	17	2.3	451	24 AAH77208	Probe #5118 for ge
64	17	2.3	570	22 ABA63465	Arthropod defensin
65	17	2.3	570	22 ABA30663	Human foetal liver
66	17	2.3	570	22 ABA30663	Probe #9129 for ge
67	17	2.3	570	22 AAK11998	Human brain expres
68	17	2.3	570	22 AAK37702	Human bone marrow
69	17	2.3	570	22 AA118460	Probe #8393 for ge
70	17	2.3	751	22 AA143577	Probe #12263 used
71	17	2.3	775	22 AAH04794	Human cDNA clone (
72	17	2.3	789	19 AAV61275	Human breast cance
73	17	2.3	789	19 AAV58509	3' cDNA sequence o
74	17	2.3	789	21 AAA06272	3' fragment of pro
75	17	2.3	789	21 AAS63480	Human immunogenic
76	17	2.3	789	22 AAS53031	Human prostate cDN
77	17	2.3	789	22 AAH93388	Human prostate tum
78	17	2.3	789	22 AAH84702	Human prostate-spe
79	17	2.3	789	22 AAH02453	Human prostate-spe
80	17	2.3	855	22 AAH34676	Prostate tumour an
81	17	2.3	876	22 AAL16182	Human colon cancer
82	17	2.3	894	21 AAC47708	Human breast cance
					Arabidopsis thalia

83	17	2.3	921	23	AA571972	DNA encoding novel	156	16	2.1	518	23	AA567091	DNA encoding novel
84	17	2.3	924	22	AA931909	Human olfactory re	c 157	16	2.1	543	22	AAH10144	Human CDNA clone (
85	17	2.3	927	22	AA542327	Human CDNA encodin	158	16	2.1	547	22	AAH11795	Human CDNA clone (
86	17	2.3	936	22	AAH45734	A thaliana Gbp-4-k	159	16	2.1	552	22	AAH76759	Human colon cancer
87	17	2.3	1045	21	AA669441	Human secreted pro	160	16	2.1	601	21	AA280058	Human colon cancer
88	17	2.3	1130	21	AA934207	Human transferrase	161	16	2.1	606	21	AAA30888	Breast cancer DDPC
89	17	2.3	1161	21	AA51047	Arabidopsis thalia	c 162	16	2.1	632	23	ABL22983	Drosophila melanog
90	17	2.3	1176	23	AA579792	DNA encoding novel	163	16	2.1	636	22	AAH68177	C glutamicum codin
91	17	2.3	1440	23	AA542927	Arabidopsis thalia	164	16	2.1	654	22	AAH04217	Human CDNA clone (
92	17	2.3	1501	21	AA681037	Human secreted pro	165	16	2.1	666	22	AAH67752	Corynebacterium gl
93	17	2.3	1583	22	AA45315	DNA encoding a GDP	c 166	16	2.1	676	22	AAH92252	Human CDNA 5'-end
94	17	2.3	1593	22	ABAI7867	Human nervous syst	c 167	16	2.1	676	22	AAH93515	Human CDNA clone r
95	17	2.3	1684	21	AA411148	Arabidopsis thalia	c 168	16	2.1	700	22	AAH92985	Human inflammatory
96	17	2.3	1722	22	AAH16413	Human CDNA sequenc	c 169	16	2.1	700	22	AAH92986	Human inflammatory
97	17	2.3	1749	23	AB105331	Drosophila melanog	c 170	16	2.1	715	23	AA568579	DNA encoding novel
98	17	2.3	1753	22	AA52878	Human CDNA encodin	171	16	2.1	717	22	AA196670	Human neuroblastom
99	17	2.3	1763	22	AA522642	Human CDNA encodin	172	16	2.1	728	7	AAH60741	Sequence encoding
100	17	2.3	1837	22	AAH81744	Human membrane ass	173	16	2.1	728	7	AAH61379	Sequence encoding
101	17	2.3	2508	21	AAH78430	Human secreted pro	174	16	2.1	728	22	AAH91877	Human CDNA 5'-end
102	17	2.3	2524	21	AA58106	Human PRO1755 nucl	175	16	2.1	728	22	AAH93283	Human CDNA clone r
103	17	2.3	2524	21	AA58106	Human PRO1755 nucl	176	16	2.1	735	20	AAH27387	Human secreted pro
104	17	2.3	2524	22	AA544449	Primer #100 used i	c 177	16	2.1	744	23	AA570603	DNA encoding novel
105	17	2.3	2527	21	AAH78398	Human secreted pro	178	16	2.1	768	22	AA158816	Human polynucleoti
106	17	2.3	2667	23	AB108223	Drosophila melanog	179	16	2.1	774	23	AA531972	Klebsiella pneumon
107	17	2.3	3060	22	AA571133	DNA encoding Dros	180	16	2.1	780	11	AAQ03647	Porcine beta FSH s
108	17	2.3	3060	23	AB113543	Drosophila melanog	c 181	16	2.1	792	22	AAH48639	Human ceramidase K
109	17	2.3	3114	23	AB126943	Drosophila melanog	c 182	16	2.1	794	23	AB115283	Drosophila melanog
110	17	2.3	3185	23	AA591627	DNA encoding novel	c 183	16	2.1	794	23	AAH14978	Trichoderma reesei
111	17	2.3	3263	22	AAH18434	Human CDNA sequenc	184	16	2.1	813	22	AAH06705	Human CDNA clone (
112	17	2.3	3472	21	AAH07588	Human piwi gene, d	185	16	2.1	816	23	AA582801	DNA encoding novel
113	17	2.3	3900	23	AA564300	DNA encoding novel	c 186	16	2.1	816	23	AAH4584	DNA encoding novel
114	17	2.3	4242	23	AB105330	Drosophila melanog	c 187	16	2.1	827	22	AAH03752	Human CDNA clone (
115	17	2.3	5114	23	AB126942	Drosophila melanog	c 188	16	2.1	831	20	AAH84475	Human secreted pro
116	17	2.3	5340	22	AA571132	DNA encoding Dros	c 189	16	2.1	831	22	ABH83258	Human secreted pro
117	17	2.3	5340	23	AB113542	Drosophila melanog	c 190	16	2.1	837	21	AAH32261	Arabidopsis thalia
118	17	2.3	7792	22	AB108222	Drosophila melanog	c 191	16	2.1	873	22	AAH92293	Bovine mammary tis
119	17	2.3	11617	22	AAH75374	Human immune/haema	c 192	16	2.1	906	22	AA197690	Human neuroblastom
120	17	2.3	11617	22	AA162937	Human genomic DNA	193	16	2.1	955	22	AA163826	Human polynucleoti
121	17	2.3	12100	22	AA540550	DNA encoding human	c 194	16	2.1	955	22	AA531620	cDNA encoding nove
122	17	2.3	12100	22	AA106438	Human reproductive	c 195	16	2.1	957	23	AA531895	Enterococcus faeca
123	17	2.3	12100	22	AAK69502	Human immune/haema	c 196	16	2.1	990	23	AA52689	E. coli DNA for ce
124	17	2.3	12100	22	AAK72028	Human immune/haema	c 197	16	2.1	1023	23	AA541009	Klebsiella pneumon
125	17	2.3	12100	22	AAK72028	Human immune/haema	c 198	16	2.1	1048	21	AAH15645	Human prostate can
126	17	2.3	15561	22	AAH7206	Human immune/haema	199	16	2.1	1061	21	AAH15520	Arabidopsis thalia
127	17	2.3	26225	22	AAH25341	Nucleotide sequenc	200	16	2.1	1063	21	AAH39111	Arabidopsis thalia
128	17	2.3	4403765	22	AA535776	Human cardiovascular	201	16	2.1	1104	22	AAH42427	Nucleotide sequenc
129	17	2.3	4411529	22	AA199683	Mycobacterium tube	202	16	2.1	1114	22	AAH91977	B thuringiensis 44
130	16	2.1	30	20	AAH34254	primer ppv.luc.Fv	203	16	2.1	1130	22	AAH91016	Human secreted pro
131	16	2.1	51	19	AAH28988	Human SNP oligonuc	204	16	2.1	1131	20	AAH84367	Human stomach carc
132	16	2.1	116	19	AAH44767	Expressed sequence	205	16	2.1	1142	24	AAH99923	Polynucleotide enc
133	16	2.1	136	20	AAH86130	Human single nucle	206	16	2.1	1147	22	AAH31687	Human polynucleoti
134	16	2.1	204	21	AAH10073	Human secreted pro	207	16	2.1	1152	18	AAH06394	B. thuringiensis 4
135	16	2.1	251	19	AAH12926	Human biallelic po	208	16	2.1	1152	21	AAH96766	DNA sequence encod
136	16	2.1	251	19	AAH11693	Human biallelic po	209	16	2.1	1152	23	AAH66918	DNA encoding novel
137	16	2.1	297	21	AA261684	cDNA encoding muri	210	16	2.1	1176	22	AAH66631	C glutamicum codin
138	16	2.1	297	22	AAH99617	Skin cell cDNA, SE	211	16	2.1	1182	22	AAH55561	C glutamicum codin
139	16	2.1	300	21	AAH00183	Human colon cancer	212	16	2.1	1197	22	AAH52325	S. epidermidis ope
140	16	2.1	323	22	AAH38293	Novel human diagno	213	16	2.1	1197	22	AAH53052	S. epidermidis ope
141	16	2.1	340	22	AAH37541	Novel human diagno	214	16	2.1	1200	24	AAH24915	Human ubiquitin-11
142	16	2.1	374	22	AAH25134	Human ovarian PCR-	c 215	16	2.1	1210	21	AAH00293	Wheat histidine bi
143	16	2.1	376	22	AAH37306	Novel human diagno	c 216	16	2.1	1215	22	AAH1566	Human polynucleoti
144	16	2.1	403	22	AAH08389	Human Ca transport	c 217	16	2.1	1222	22	AAH34885	Human colon cancer
145	16	2.1	447	22	AAH29547	Drosophila melanog	218	16	2.1	1228	23	AAH66650	DNA encoding novel
146	16	2.1	449	21	AAH08956	Human secreted pro	219	16	2.1	1233	22	AAH71044	C. glutamicum SRT
147	16	2.1	477	22	AAH75723	Human immune/haema	220	16	2.1	1241	22	AAH91952	B thuringiensis 44
148	16	2.1	486	22	AAH71041	C. glutamicum SRT	221	16	2.1	1249	21	AAH59169	Human secreted pro
149	16	2.1	494	24	AB199746	Mouse ischaemic co	c 222	16	2.1	1256	23	AB114665	Drosophila melanog
150	16	2.1	498	22	AAH83780	Human ovarian tumo	c 223	16	2.1	1276	22	AAH94548	Human full-length
151	16	2.1	502	22	AAH13322	Human cDNA clone (c 224	16	2.1	1322	22	AAH8608	Human polynucleoti
152	16	2.1	506	22	AAH62150	Human foetal liver	225	16	2.1	1335	23	AAH1592	DNA encoding novel
153	16	2.1	506	22	AAH10469	Human brain expres	226	16	2.1	1340	21	AAH9090	Human pancreatic c
154	16	2.1	506	22	AAH36370	Human bone marrow	c 227	16	2.1	1356	22	AAH00414	Human calcium ion
155	16	2.1	506	22	AAH42101	Probe #10787 used	228	16	2.1	1376	22	AAH08155	Mammalian toxicolo

229 16 2.1 1390 20 AAV84524 Human secreted pro
230 16 2.1 1390 22 ABA83307 drosophila melanog
231 16 2.1 1411 23 ABL18883 Partial mouse DNAX
232 16 2.1 1431 20 AAZ08868 Human immune/haema
233 16 2.1 1431 22 AAK66183 Human immune/haema
234 16 2.1 1432 22 AAK66184 DNA encoding novel
235 16 2.1 1436 23 AAS93208 Porcine adrenomedu
236 16 2.1 1493 22 AAF27229 Porcine adrenomedu
237 16 2.1 1493 22 AAF29139 Human alpha nicoti
238 16 2.1 1501 22 AAS01280 Human ATP-dependen
239 16 2.1 1509 22 AAI71301 Human ORF2395
240 16 2.1 1515 21 AAC76840 B thuringiensis 58
241 16 2.1 1521 22 AAF92018 Drosophila melanog
242 16 2.1 1522 23 ABL07681 Human prostate cDN
243 16 2.1 1533 22 AAS64173 P768P full length
244 16 2.1 1533 22 AAH93937 Human prostate cDN
245 16 2.1 1536 22 AAS64172 P768P full length
246 16 2.1 1536 22 AAH93936 Human protein enco
247 16 2.1 1543 22 AAH99635 Human stomach carc
248 16 2.1 1544 22 ABA09206 Human polynucleoti
249 16 2.1 1544 22 AAK52671 Human stomach canc
250 16 2.1 1570 22 AAT93848

ALIGNMENTS

RESULT 1
ID AAV37699 standard; DNA: 824 BP.
XX AAV37699;
AC AAV37699;
DT 18-SEP-1998 (first entry)
DE Human parotid secretory protein (HPSP) encoding DNA.
XX
XX Parotid secretory protein; human; cancer; autoimmune disease;
KW secretory tissue; gastrointestinal tissue; HPSP; Sjorgen's syndrome;
KW Graves disease; thyroiditis; insulin-dependent diabetes; pancreatitis;
KW ulcerative colitis; Crohn's disease; atrophic gastritis; ss.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT CDS 48..797
FT /*tag= a
FT /product= "parotid secretory protein"
FT /note= "#"
XX
XX W09821329-A1.
XX
XX 22-MAY-1998.
XX
XX 07-NOV-1997; 97WO-US20651.
XX
XX 14-NOV-1996; 96US-0749288.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Bandman O, Goli SK;
XX
XX WPI: 1998-297933/26.
XX P-PSDB; AAW60682.
XX
XX New parotid secretory protein - useful for, e.g. treatment of cancer
PT and auto-immune disease, particularly of secretory or
PT gastrointestinal tissues
XX
XX Claim 5; Fig 1A-C; 65pp; English.
XX
XX This DNA encodes a human parotid secretory protein (HPSP). Antagonists
CC that bind specifically to, and modulate activity of HPSP are used to

CC treat cancer and autoimmune diseases particularly of secretory or
CC gastrointestinal tissue, e.g. cancer of salivary gland, thyroid,
CC prostate, breast, gastrointestinal tract or pancreas, Sjorgen's syndrome,
CC Graves disease, thyroiditis, insulin-dependent diabetes, pancreatitis,
CC ulcerative colitis, Crohn's disease and atrophic gastritis. Host cells
CC containing expression vectors comprising the HPSP nucleic acid are used
CC to produce recombinant HPSP which is used to generate antibodies and to
CC screen for its antagonists. Antibodies are useful directly as
CC antagonists, to transport drugs to HPSP-expressing cells, to detect cells
CC that express HPSP, to monitor patients being treated with HPSP, and for
CC purification of HPSP from natural sources. Expression of HPSP may
CC indicate cell proliferation. HPSP nucleic acid or its fragments are used
CC to detect HPSP-encoding sequences (optionally after amplification by PCR)
CC by hybridisation, particularly for diagnosis and monitoring of disease,
CC but also for mapping the chromosomal sequence.
XX
XX Sequence 824 BP: 226 A: 225 C: 184 G: 189 T: 0 other;

Query Match 100.0%; Score 746; DB 19; Length 824;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTTCAGCTTGGAACTTGTCTCTGCGGGCTGCTCACTGGGACCTCAGAGTC 60
DB 47 gatcttcagctttggaaacttgttctctgtcggggtgctcactgggacctcagagtc 106
QY 61 TCTTCTTGACAATCTTGGCAATGACCTAAGCAATGTCTGGGTAAGCTGGAACCTCTTCT 120
DB 107 tcttcttgacaatcttggcaatgacctaaagcaatgctggtgaagctggaacctgtct 166
QY 121 TCAGSAGGACTTGAGACAGTGTGACAACTACTCTTAAAGGCATCCTTGAGAACTGAAGT 180
DB 167 tcacggaggacttgagacagtgtgacaatactcttaagcatctcttgaaacctgaaggt 226
QY 181 CGACTAGGAGTGTCTCAGAAATCCAGTGTCTGGCAACTGGCCCAAGCAGAGCCCGCA 240
DB 227 cgactaggagtgctcagaaatccagtgtctggcaactggcccaagcagaagcccgagga 286
QY 241 AGCTGAGAAATGCTGCAAAATGTCTTCTAGCTGCTTCCAACTAAACACGGACATTTT 300
DB 287 agctgagaaatgctgcaaaatgtcttcttaagctgtctccaaactaaacacgagacattt 346
QY 301 TGGGTTGAAAATCAGCAACTCCCTCATCTCCCTGGATGTCAAAGCTGAACCCGATGATGG 360
DB 347 tgggttgaatacagcaactccctcatctcctggatgtcaaaagctgaaccgatgatgg 406
QY 351 CAAGGGCTTAACTGAGCTTCCCTGTACCGGGAATGTCTACTGTGCGCGGCGCCATCAT 420
DB 407 caaaggccttaacctgagcttccctgtcaaccggaatgtcactgtgcccgggcccacat 466
QY 421 TGGCCAGATTATCAACCTGAAAGCTCTCTTGGACCTCTGACCGCAGTCACAATTGAAAC 480
DB 467 tggccagattataaactgaagcctccttgagacctcctgacccgagtcacaattgaac 526
QY 481 TGATCCCCAGACACACAGCGCTGTTCCTGCTCTGGGAGATGGCCGAGTGACCCACCAG 540
DB 527 tgatccccagacacacagcgcctgttgcgtctctgggagaatgcgccagtgacccaaccag 586
QY 541 CATCTCACTTTCCTTGGTGGACAAACACAGCCAAATCATCAACAAGTTCGTGAATAGCGT 600
DB 587 catctcaacttctctgtggacaaaacacagccaaatcatcaacaagtctgtaagcgt 646
QY 601 GATCAACACGCTGAAAAGCACTGTATCTCCCTGCTGCGAAGAGATATGTCCACTGAT 660
DB 647 gatacaacacgctgaaaagcactgtatctctcctgctgcagaagagatgtgcactgat 706
QY 661 CCGCATCTTTCATCCACTCCCTGGATGTGAATGTCAATTCAGCAGGTGCTCGATAATCTCA 720
DB 707 ccgcatcttctacactccctggaatgtgaatgtcattcgaaggtctgcgataatctca 766
QY 721 GCACAAAACCCAGCTGCAAAACCCCTCA 746
DB 746

Db 767 gcacaaacccagctgcaaacctca 792

RESULT 2
AAV44759

ID AAV44759 standard; DNA: 1028 BP.

XX
AC AAV44759;

DT 16-OCT-1998 (first entry)

XX
XX Human parotid secretory protein coding sequence.

XX
XX Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
XX non-immune defensive disorder; immune system disorder; cancer; human;
XX therapy; diagnosis; ss.

XX
OS Homo sapiens.

XX
XX Key Location/Qualifiers
XX CDS 49..798
XX /*tag= a
XX /*product= hPSP
XX sig_peptide 49..102
XX /*tag= b
XX mat_peptide 103..795
XX /*tag= c

XX
XX WO9828420-A1.

XX
XX 02-JUL-1998.

XX
XX 18-DEC-1997; 97WO-US23522.

XX
XX 23-DEC-1996; 96US-0034429.

XX
XX (HUMA-) HUMAN GENOME SCI INC.

XX
XX Duan R, Ruben SM;

XX
XX WPI: 1998-377651/32.

XX
XX P-PSDB: AAW69221.

XX
XX New nucleic acid encoding human parotid secretory protein or its
XX fragments - useful for diagnosis and treatment of, e.g. digestive
XX and endocrine disorders and for drug screening

XX
XX Claim 2; Fig 1; 94pp; English.

XX
XX This sequence encodes the human parotid secretory protein (hPSP) of the
XX invention. The hPSP DNA is useful for chromosome identification and
XX isolation of the corresponding genomic DNA. The DNA and protein can be
XX used to detect abnormal levels of hPSP (in standard blotting, or
XX amplification or immuno assays), particularly for diagnosis of digestive,
XX non-immune defensive, endocrine or immune system disorders. A particular
XX application is diagnosis of cancers of the salivary gland, thymus and
XX pancreas which are associated with high levels of hPSP. The protein is
XX also useful as antifungal, antibacterial, antiparasitic and antiviral
XX agents and may be expressed in vivo from the DNA. The protein, or cells
XX expressing it, are used in screening tests to identify specific
XX (ant)agonists, e.g. antibodies (Ab), polypeptides and antisense nucleic
XX acids, which are potentially useful for treating conditions associated
XX with excessive hPSP production. Cells containing the DNA are used to
XX express the recombinant protein and this can be used to raise Ab, useful
XX for diagnosis, therapy, for affinity purification and to identify
XX hPSP-binding proteins.

XX
XX Sequence 1028 BP; 274 A; 299 C; 222 G; 233 T; 0 other;


```
Db 186 tcaagagagacttgagacagttgacaatactcttaagagcatcttggagaaactgaaggt 245
Qy 181 CGACCTAGAGTGTCTTGAAGATCCAGTGTGCTGCACTGGCCAAACAGAGCCCGCAGGA 240
Db 246 cgacctagagtgcttcagaaatccagtgcttgcaactgcccagcagcaagcccgagga 305
Qy 241 AGCTCAGAAATGTGTGAACATGTCTTCTTAAGCTCTCTTCCAACTAATACAGGACATTTT 300
Db 306 agctagaaaattgtcgaacaatgcaattcttaagctcttccaaactaagcaggaactttt 365
Qy 301 TGGTTGAAATCAGCACTCCCTCTCTGCGATGTCAAGCTGAACCGATGATGAG 360
Db 366 tgggttgaaaatcagcaactccctccctcttgatgctgcaagctgaaacgctgatgatgg 425
Qy 361 CAAAGCCCTTAACCTGAGCTTCCCTGTACACCGCAATGTCACTGTGCGCGGCCCATCAT 420
Db 426 caaagccttaactgaactccctctctcgcgcgaatgtcaactgtgscgggccccatcat 485
Qy 421 TGGCAGATTTATCAACCTGAAGCTCTCTTGGACCTCTGACCCGAGTCACAATTGAAAC 480
Db 486 tggccagattatcaacctgaaagctctcttgacacctctgacgcagctcaacaattgaaac 545
Qy 481 TCATCCCCAGACACACACAGCTGTGCGTCTCTGGGAGATGCGCCAGTGACCCAAACAG 540
Db 546 tgatccccagacacacacagcctgtgctgctgaggaatgcccagtgaccccaacag 605
Qy 541 CATCTACTTTCTCTGTGTCGACACACAGCCAAATCATCAACAGTTCTGTGAATAGCGT 600
Db 606 catctcactcttctgtggaacacacagcccaaatcatcaacaagtctgtgaatagcgt 665
Qy 601 GATCAACAGCTGAAACGACTGTATCTCTCTCTGTCAGAGGAGATATGTCCACTGAT 660
Db 666 gatcacacgcctgaaaagcactgtatctctctctctgctgcagaagagagatagtccactgat 725
Qy 661 CGCATCTTTATPCCACTCCCTCGGATGTGAATGTGATTCAGCAGGTCTGCAATATCTCA 720
Db 726 cgcgatcttctcactcctctctggaatgtgaatgtcattcagcaggtctgcataatctca 785
Qy 721 GCACAAACCCAGCTGCAAAACCCCTCA 746
Db 786 gcacaaacccagctgcacacccctca 811

RESULT 6
AAH99686
ID AAH99686 standard; cDNA; 1041 BP.
XX
AC AAH99686;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:521.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antilagerant; haemostatic; vulnerary; antilulcer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.
XX
OS Homo sapiens.
XX
XX WO200153455-A2.
PN
```

```
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000WO-US35017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-457603/49.
DR P-PSDB; AAM25745.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
PS Claim 1; Page 579; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antilagerant; haemostatic; vulnerary;
CC antilulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antitense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
XX
SQ Sequence 1041 BP; 284 A; 299 C; 227 G; 231 T; 0 other;
```

```
Query Match 54.2%; Score 404; DB 22; Length 1041;
Best Local Similarity 100.0%; Pred. No. 2e-194;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATGCTTTCAGCTTGGAAACTTGTCTCTCTGTCGCGGCTGCTCACTGGGACCTCAGAGTC 60
Db 52 gatgcttcagcttggaaactgttctctctctgctgctgctgctgctgctgctgctgctgct 111

Qy 61 TCTTCTTGCAATCTTGGCAATGACCTAAGCAATGCTGCTGGATAGCTGGAACCTGTCT 120
Db 112 tcttcttgcaatcttggcaatgacctgaagcaatgctgctgataagctggaacctgtctct 171

Qy 121 TCACGAGGACCTTGACACAGTTGACAACTACTCTTTAAAGGCATCTCTTGAGAACTGAAGGT 180
Db 172 tcacgaggaccttgacacagttgacaaactctttaaaggcatctcttgagaaactgaaggt 231

Qy 181 CGACCTAGGAGTCTTTCAGAAATCCAGTGTGTCGCACTGGCCAGCAGAGGCCAGGA 240
Db 232 cgacctagagtgctgtcagaaatccagtgcttggcaactgcccagcaggaagccccagga 291

Qy 241 AGCTCAGAAATGTGTGAACATGTCTTCTTAAGCTCTCTTCCAACTAATACAGGACATTT 300
Db 292 agctcagaaatgtgtgaacaatgtcttcttcttaagctgtcttccaaactaacacggacattt 351

Qy 301 TGGTTGAAATCAGCACTCCCTCTCTGCGATGTCAAGCTGAACCGATGATGAG 360
Db 352 tgggttgaaaatcagcaactccctccctctgctgctgctgctgctgctgctgctgctgctgct 411
```

QY 361 CAAAGGCGCTTAACCTGAGCTGCCCTGTCACCGCGAATGTCAGT 404
|||||
Db 412 caaaggccttaacctgagcttccctgtcaccggaagtgcactg 455

RESULT 7

AAV44761
ID AAV44761 standard; DNA: 538 BP.
XX
AC AAV44761;
XX
DT 16-OCT-1998 (first entry)
XX
DE Expressed sequence tag HSGSCL3R.

XX Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
KW non-immune defensive disorder; immune system disorder; cancer; human;
KW therapy; diagnosis: ss.
XX
OS Homo sapiens.

XX WO9828420-A1.
XX
XX 02-JUL-1998.
XX
XX 18-DEC-1997; 97WO-US233522.
XX
XX 23-DEC-1996; 96US-0034429.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Duan R, Ruben SM;
XX
XX WPI; 1998-377651/32.

XX New nucleic acid encoding human parotid secretory protein or its
PT fragments - useful for diagnosis and treatment of, e.g. digestive
PT and endocrine disorders and for drug screening

XX Disclosure; Page 76; 94pp; English.

XX This sequence represents an expressed sequence tag with homology to the
CC DNA of the invention, and is all so specifically stated as not being
CC contained within the DNA of the invention. The DNA of the invention
CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
CC useful for chromosome identification and isolation of the corresponding
CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
CC hPSP (in standard blotting, amplification or immuno assays), particularly
CC for diagnosis of digestive, non-immune defensive, endocrine or immune
CC system disorders. A particular application is diagnosis of cancers of the
CC salivary gland, thymus and pancreas which are associated with high levels
CC of hPSP. The protein is also useful as antifungal, antibacterial,
CC antiparasitic and antiviral agents and may be expressed in vivo from the
CC DNA. The protein, or cells expressing it, are used in screening tests to
CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
CC antisense nucleic acids, which are potentially useful for treating
CC conditions associated with excessive hPSP production. Cells containing
CC the DNA are used to express the recombinant protein and this can be used
CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
XX to identify hPSP-binding proteins.

XX Sequence 538 BP; 140 A; 126 C; 128 G; 124 T; 20 other;

Query Match 29.2%; Score 218; DB 19; Length 538;
Best Local Similarity 100.0%; Pred. No. 3.2e-100;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGCTTTGGAACCTGTCTCTCTGTCGGCGCTGCTCACTGGGACCTCAGAGTCTCTCT 67
|||||

Db 45 cagctttggaacctgtctctctgtgctgctcactgggacctcagagtcctctctt 104

QY 68 GACAATCTTGGCAATGACCTAAGCAATGCTGTGGATAAGCTGGAACCTGTTCTTCACGAG 127
|||||
Db 105 gacaatcttggcaatgacctaaagcaatgctgtgataagctggaacctgtcttcacgag 164
|||||
QY 128 GGACTTGGAGACAGTTGACAATCTCTTAAGGCATCTTTGAGAAACTGAAGTCGACCTA 187
|||||
Db 165 ggacttggagacagttgacaataactcttaaggcatctcttgagaaaactgaagtcgacct 224
|||||

QY 188 GGAGTGCTTCCAGAAATCCAGTGGTGGCAACTGGCCAA 225
|||||
Db 225 ggagtgcttccagaatccagtgcttggcaactggccaa 262

RESULT 8

AAV44764
ID AAV44764 standard; DNA: 406 BP.

XX AAV44764;

XX 16-OCT-1998 (first entry)

XX Expressed sequence tag HSGSC78R.

XX Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
KW non-immune defensive disorder; immune system disorder; cancer; human;
KW therapy; diagnosis: ss.

XX Homo sapiens.

XX WO9828420-A1.

XX 02-JUL-1998.

XX 18-DEC-1997; 97WO-US233522.

XX 23-DEC-1996; 96US-0034429.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Duan R, Ruben SM;

XX WPI; 1998-377651/32.

XX New nucleic acid encoding human parotid secretory protein or its
PT fragments - useful for diagnosis and treatment of, e.g. digestive
PT and endocrine disorders and for drug screening

XX Disclosure; Page 77; 94pp; English.

XX This sequence represents an expressed sequence tag with homology to the
CC DNA of the invention, and is all so specifically stated as not being
CC contained within the DNA of the invention. The DNA of the invention
CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
CC useful for chromosome identification and isolation of the corresponding
CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
CC hPSP (in standard blotting, amplification or immuno assays), particularly
CC for diagnosis of digestive, non-immune defensive, endocrine or immune
CC system disorders. A particular application is diagnosis of cancers of the
CC salivary gland, thymus and pancreas which are associated with high levels
CC of hPSP. The protein is also useful as antifungal, antibacterial,
CC antiparasitic and antiviral agents and may be expressed in vivo from the
CC DNA. The protein, or cells expressing it, are used in screening tests to
CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
CC antisense nucleic acids, which are potentially useful for treating
CC conditions associated with excessive hPSP production. Cells containing
CC the DNA are used to express the recombinant protein and this can be used
CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
XX to identify hPSP-binding proteins.

XX Sequence 406 BP; 110 A; 95 C; 95 G; 100 T; 6 other;

Query Match 28.3%; Score 211; DB 19; Length 406;

Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 CTTTGGAAACTGTTCTCTGTCGGGGCTGCTCAGTCGACCTCAGAGTCTTCTTGTGAC 70
|||||
Db 49 ctttggaaactgtctctctgtgctgctgctcactgagctcagctcctctcttgac 108
|||||
QY 71 AATCTTGCAATGACCTAAGCAATGTCGTGGATAAGCTGGAACCTGTTCTTCACGAGGGA 130
|||||
Db 109 aatcttggcaatgacctaaagcaatgtctgataagctggaacctgctcttcacgagga 168
|||||
QY 131 CTTGAGACAGTTGACAACTACTCTTAAGGATCTCTGAGAACTGAAGTGTGACCTAGGA 190
|||||
Db 169 cttgagacagttgacaataactctttaaaggcatctcttgagaaactggaaggtgcacotagga 228
|||||
QY 191 GTGCTTCAGAAATCCAGTGTGTCGCAACTGG 221
|||||
Db 229 gtgcttcagaaatccagtgcttggaactgg 259
|||||

RESULT 9
AAV44765
ID AAV44765 standard; DNA: 493 BP.
AC AAV44765;
XX
XX
DT 16-OCT-1998 (first entry)
DE
DE Expressed sequence tag HSPMD36R.
KW Parotid secretory protein; hpsp; digestive disorder; endocrine disorder;
KW non-immune defensive disorder; immune system disorder; cancer; human;
KW therapy; diagnosis; ss.
XX
XX Homo sapiens.
XX
XX WO9828420-A1.
XX
PD 02-JUL-1998.
XX
XX
PF 18-DEC-1997; 97WO-US23522.
XX
XX
PR 23-DEC-1996; 96US-0034429.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Duan R, Ruben SM;
XX
XX WPI; 1998-377651/32.
XX
XX New nucleic acid encoding human parotid secretory protein or its
PT fragments - useful for diagnosis and treatment of, e.g. digestive
PT and endocrine disorders and for drug screening
XX
XX Disclosure; Page 78; 94pp; English.

This sequence represents an expressed sequence tag with homology to the
CC DNA of the invention, and is all so specifically stated as not being
CC contained within the DNA of the invention. The DNA of the invention
CC encodes the human parotid secretory protein (hpsp). The hpsp DNA is
CC useful for chromosome identification and isolation of the corresponding
CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
CC hpsp (in standard blotting, amplification or immuno assays), particularly
CC for diagnosis of digestive, non-immune defensive, endocrine or immune
CC system disorders. A particular application is diagnosis of cancers of the
CC salivary gland, thymus and pancreas which are associated with high levels
CC of hpsp. The protein is also useful as antifungal, antibacterial,
CC antiparasitic and antiviral agents and may be expressed in vivo from the
CC DNA. The protein, or cells expressing it, are used in screening tests to
CC identify specific (ant)agonists e.g. antibodies (Ab), polypeptides and
CC antisense nucleic acids, which are potentially useful for treating
CC conditions associated with excessive hpsp production. Cells containing
CC the DNA are used to express the recombinant protein and this can be used

to raise Ab, useful for diagnosis, therapy, for affinity purification and
CC to identify hpsp-binding proteins.
XX
SQ Sequence 493 BP; 118 A; 143 C; 107 G; 122 T; 3 other;

Query Match 24.4%; Score 182; DB 19; Length 493;
Best Local Similarity 100.0%; Pred. No. 5.5e-82;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 555 ACACAGCCAAATCATCAACACTTCTGTAATAGCGTGCATCAACAGCGCTGAAAGCACTGT 624
|||||
Db 93 acacagccaaatcatcaaacagttcgtgaaatagcgtgatacaacacgctgaaagcactgt 152
|||||
QY 625 ATCTCTCCCTGCTGGCAGAGGAGATATGTCACACTGATCCGGCATCTTCATCCACTCCCTGGA 684
|||||
Db 153 atctctccctgctgcagaagagatattgccactgatcgcgatcttcactccctgga 212
|||||
QY 685 TGTGAATGTCATTGACGAGGTGCTGCGATAATCTTCAGCAGCAAAACCCAGCTGCAACCCCT 744
|||||
Db 213 tgtgaatgtcttcagcaggtcgtcgataatctcagcaaaaacccagctgcaaacct 272
|||||
QY 745 CA 746
II
Db 273 ca 274
II

RESULT 10
AAC23485
ID AAC23485 standard; cDNA: 317 BP.
XX
AC AAC23485;
XX
DT 06-OCT-2000 (first entry)
DE
DE Human secreted protein 5' EST, SEQ ID NO: 27560.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EPI033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 27560; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX Sequence 317 BP: 82 A; 100 C; 68 G; 67 T; 0 other;
SQ

Query Match 23.7%; Score 177; DB 21; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.9e-79;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 570 GCACAAATCATCAACAAAGTTCGTGAATGACGCTGATCAACACGCTGAAAGCACTGTATCCT 629
DB 1 gccaaatcatcaacaagtctgtgaatgactgacacacgctgaaaagcaactgtatcct 60
QY 630 CCTCTCTCAGAACGAGATATGTCACCTGATCCGCTCTTTCATCCACTCCCTCGATGTGA 689
DB 61 cccctctcagaaggagatagtcacactgatcccgatccatcttcacactccctggatgtga 120
QY 690 ATGTCATTACGAGGCTGCTGATATCTCAGCACAAACCCAGCTGCACACCCCTCA 746
DB 121 atgtcattcagcagctgcgataatctcagacaaacccagctgcaaacctca 177

RESULT 11
AAV44760
ID AAV44760 standard; DNA: 449 BP.
AC AAV44760;
DT 16-OCT-1998 (first entry)
DE Expressed sequence tag HSCSA61R.
XX
XX Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
KW non-immune defensive disorder; immune system disorder; cancer; human;
KW therapy; diagnosis: ss.
XX Homo sapiens.
XX
XX WO9828420-A1.
PN
PD 02-JUL-1998.
XX
XX 18-DEC-1997; 97WO-US23522.
XX
XX 23-DEC-1996; 96US-0034429.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Duan R, Ruben SM;
PI
XX
XX WPI; 1998-377651/32.
DR
XX
XX New nucleic acid encoding human parotid secretory protein or its
PT fragments - useful for diagnosis and treatment of, e.g. digestive
PT and endocrine disorders and for drug screening
XX
XX Disclosure: Page 75-76; 94pp; English.
PS
XX This sequence represents an expressed sequence tag with homology to the
CC DNA of the invention, and is all so specifically stated as not being
CC contained within the DNA of the invention. The DNA of the invention
CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
CC useful for chromosome identification and isolation of the corresponding
CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
CC hPSP (in standard blotting, amplification or immuno assays), particularly
CC for diagnosis of digestive, non-immune defensive, endocrine or immune
CC system disorders. A particular application is diagnosis of cancers of the
CC salivary gland, thymus and pancreas which are associated with high levels
CC of hPSP. The protein is also useful as antifungal, antibacterial,
CC antiparasitic and antiviral agents and may be expressed in vivo from the
CC DNA. The protein, or cells expressing it, are used in screening tests to

CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
CC antisenese nucleic acids, which are potentially useful for treating
CC conditions associated with excessive hsp production. Cells containing
CC the DNA are used to express the recombinant protein and this can be used
CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
CC to identify hPSP-binding proteins.
XX
SQ Sequence 449 BP: 122 A; 107 C; 106 G; 110 T; 4 other;
Query Match 21.3%; Score 159; DB 19; Length 449;
Best Local Similarity 100.0%; Pred. No. 2.4e-70;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGCTTCAGCTTTGGAAACTGTCTCTCTGCGGGCTGCTCACTGGACCTCAGAGTC 60
DB 50 gatgtcttcagctttggaaactgttctctctgctcactgctcactggacctcagagtc 109
QY 61 TCTTCTTGACATCTTGCAATGACCTAAGCAATGTCTGTGATTAAGCTGGAACCTGTCT 120
DB 110 tcttcttgacatcttgcaatgaccttaagcaatgtcgtggataagctggaacctgttct 169
QY 121 TCACGAGGAGCTTGACAGCTGACATACTCTTTAAAGG 159
DB 170 tcacgaggagacttgagacagttgacaatactcttaagg 208

RESULT 12
AAK89798
ID AAK89798 standard; DNA: 7524 BP.
AC AAK89798;
DT 05-NOV-2001 (first entry)
XX
XX Human digestive system antigen genomic sequence SEQ ID NO: 3374.
DE Human digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
XX Homo sapiens.
X
X WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01324.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.

```
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-02311242.
PR 08-SEP-2000; 2000US-02311243.
PR 08-SEP-2000; 2000US-02311244.
PR 08-SEP-2000; 2000US-02311413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234927.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0245475.
PR 08-NOV-2000; 2000US-0245477.
PR 08-NOV-2000; 2000US-0245478.
PR 08-NOV-2000; 2000US-0245523.
PR 08-NOV-2000; 2000US-0245524.
PR 08-NOV-2000; 2000US-0245525.
PR 08-NOV-2000; 2000US-0245526.

PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
XX
XX Polynucleotides encoding digestive system antigens, useful for
XX diagnosing, treating, preventing and/or prognosing disorders of the
XX digestive system, particularly cancer and cancer metastases -
XX
XX Disclosure: SEQ ID NO 3374; 986pp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human digestive system antigens. These can be used in the
XX diagnosis, treatment and prevention of digestive system disorders,
XX including cancer, Meckel's diverticulum, bacterial or parasitic
XX infections, appendicitis, Hirschsprung's disease, chronic colitis or
XX ulcerative colitis. The present sequence is a genomic DNA fragment
XX encoding a digestive system antigen of the invention.
XX
XX Sequence 7524 BP; 2170 A; 1586 C; 1582 G; 2186 T; 0 other;
XX

Query Match 21.3%; Score 159; DB 22; Length 7524;
Best Local Similarity 100.0%; Pred. No. 2.4e-70;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGCTTCAGCTTGAACCTGTTCTCTGTGGCGCTGCTCACTGGGACCTCAGAGTC 60
Db 977 gatgcttcagcttggaaactgttctctgtgcyggtgctgctggaacttcagagtc 1036
QY 61 TCTTCTTGACAATCTTGCGAATGACCTAAGCAATGCTCGATGAGCTGGAACCTGTCT 120
|||||
```

Db 1037 tcttctgacaattctggcaatgacctaagcaatgtcgtgataagctggaacctgtct 1096

QY 121 TCACGAGGACTTGAGACAGTTGACATACTCTTAAAGG 159
|||||

Db 1097 tcacgaggaacttgagacagtgatgacaatactctttaaagg 1135
|||||

RESULT 13
ABA60313
ID ABA60313 standard; DNA: 581 BP.
XX
AC ABA60313;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #8618.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
DR Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX
PT
PS Claim 1: SEQ ID NO 8618; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 581 BP; 142 A; 161 C; 127 G; 151 T; 0 other;

Query Match 20.6%; Score 154; DB 22; Length 581;
Best Local Similarity 100.0%; Pred. No. 8.3e-68;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 GGCCTCATCTGGCCAGATTATCAACCTGAAAGCCTCTCTGACCTGACCGCAGTCA 470
|||||

Db 81 ggccctcatctggccagattatcaacctgaaagcctctcttgacctctgaccgcagtca 140
|||||

QY 471 CAATTGAAACTGATCCCGACAGACACACACAGCCTGTTCGCTCTGGGAGAAATGCGCCAGTG 530
|||||

Db 141 caattgaaactgatcccgacagacacacacagcctgttgcgcgtctctggagaaatgcgccagt 200
|||||

QY 531 ACCCAACAGCATCTCACTTTCTCTTGTGGACAA 564
|||||

RESULT 15

Db 201 acccaaccagcatctcaactttctcttctgtgagaaa 234

RESULT 14
AAK0593
ID AAK0593 standard; DNA: 581 BP.
XX
AC AAK0593;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 8584.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
DR Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
PT
PS Example 4: SEQ ID NO: 8584; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
CC
SQ Sequence 581 BP; 142 A; 161 C; 127 G; 151 T; 0 other;

Query Match 20.6%; Score 154; DB 22; Length 581;
Best Local Similarity 100.0%; Pred. No. 8.3e-68;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 GGCCTCATCTGGCCAGATTATCAACCTGAAAGCCTCTCTGACCTGACCGCAGTCA 470
|||||

Db 81 ggccctcatctggccagattatcaacctgaaagcctctcttgacctctgaccgcagtca 140
|||||

QY 471 CAATTGAAACTGATCCCGACAGACACACACAGCCTGTTCGCTCTGGGAGAAATGCGCCAGTG 530
|||||

Db 141 caattgaaactgatcccgacagacacacacagcctgttgcgcgtctctggagaaatgcgccagt 200
|||||

QY 531 ACCCAACAGCATCTCACTTTCTCTTGTGGACAA 564
|||||

Db 201 acccaaccagcatctcaactttctcttctgtgagaaa 234


```
PN WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00667.
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX Claim 4: SEQ ID NO 21157; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 153 BP; 38 A; 53 C; 30 G; 32 T; 0 other;
SQ
Query Match 20.5%; Score 153; DB 22; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.7e-67;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 412 GCCCATCATTTGCCAGATTATCAACCTCAAGCCTCTTGGACCTCTTGACCGCAGTCAC 471
Db 1 gccatcattggccagattatcaacctgaagcctcttggacctcttgaccctccgaccgagtcac 60
QY 472 AATTGAAGTATGATCCCGACACACACACACACACCTCTTGGCGGAGAAATCGGCCAGTGA 531
Db 61 aattgaactgatcccgacacacacacacacacagcctgttgcgtctctggagaatgagccatga 120
QY 532 CCCAACCCAGCATCTCAGTTCTCTTCTGAGAA 564
Db 121 cccaaccagcatctcactttcttctgtggacaa 153
RESULT 18
AAK21285
ID AAK21285 standard; DNA; 153 BP.
XX AAK21285;
XX 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe SEQ ID NO: 21276.
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX Homo sapiens.
XX WO200157275-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00668.
XX 04-FEB-2000; 2000US-0180312.
PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00667.
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT Brains -
XX Example 4: SEQ ID NO: 21276; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX Sequence 153 BP; 38 A; 53 C; 30 G; 32 T; 0 other;
SQ
Query Match 20.5%; Score 153; DB 22; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.7e-67;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 412 GCCCATCATTTGCCAGATTATCAACCTCAAGCCTCTTGGACCTCTTGACCGCAGTCAC 471
Db 1 gccatcattggccagattatcaacctgaagcctcttggacctcttgaccctccgaccgagtcac 60
QY 472 AATTGAAGTATGATCCCGACACACACACACACCTCTTGGCGGAGAAATCGGCCAGTGA 531
Db 61 aattgaactgatcccgacacacacacacacacagcctgttgcgtctctggagaatgagccatga 120
QY 532 CCCAACCCAGCATCTCAGTTCTCTTCTGAGAA 564
Db 121 cccaaccagcatctcactttcttctgtggacaa 153
RESULT 19
AAK47441
ID AAK47441 standard; DNA; 153 BP.
XX AAK47441;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed single exon probe SEQ ID NO: 21998.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00668.
XX 04-FEB-2000; 2000US-0180312.
```



```

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000US-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 21998; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
SQ Sequence 153 BP; 38 A; 53 C; 30 G; 32 T; 0 other;

Query Match          20.5%; Score 153; DB 22; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.7e-67;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 GCCCATCATTTGGCCGAGATTATCAACCTGAAAGCCTCTTGGACCTCTGACCGCAGTCA 471
DB 1 gcccatcattggccgagattatcaacctgaagcctcttggacctctgaccgcagtcac 60
QY 472 AATTGAAACTGATCCCGACACACACAGCCTGTTCCTGCTGGGAGATGCGCCAGTGA 531
DB 61 aattgaaactgatcccgacacacacagcctgttcctgctgggagaatgcgcagtgga 120
QY 532 CCCAACCCAGCATCTCACTTCTTCCTTGGGACAA 564
DB 121 cccaaccagcatctcaacttcttctgtggaaca 153

RESULT 20
AAI53276
ID AAI53276 standard; DNA; 153 BP.
XX
XX AAI53276;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #21962 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 03-JUN-2000; 2000US-0608408.
XX
XX 21-SEP-2000; 2000US-0632366.
XX
XX 27-SEP-2000; 2000US-0234687.
XX
XX 04-OCT-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000US-0024263.

```

```

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 21962; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 153 BP; 38 A; 53 C; 30 G; 32 T; 0 other;

Query Match          20.5%; Score 153; DB 22; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.7e-67;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 GCCCATCATTTGGCCGAGATTATCAACCTGAAAGCCTCTTGGACCTCTGACCGCAGTCA 471
DB 1 gcccatcattggccgagattatcaacctgaagcctcttggacctctgaccgcagtcac 60
QY 472 AATTGAAACTGATCCCGACACACACAGCCTGTTCCTGCTGGGAGATGCGCCAGTGA 531
DB 61 aattgaaactgatcccgacacacacagcctgttcctgctgggagaatgcgcagtgga 120
QY 532 CCCAACCCAGCATCTCACTTCTTCCTTGGGACAA 564
DB 121 cccaaccagcatctcaacttcttctgtggaaca 153

RESULT 21
AAV44763
ID AAV44763 standard; DNA; 374 BP.
XX
XX AAV44763;
XX
XX 16-OCT-1998 (first entry)
XX
XX Expressed sequence tag HSPA114K.
XX
XX Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
XX non-immune defensive disorder; immune system disorder; cancer; human;
XX therapy; diagnosis; ss.
XX
XX Homo sapiens.
XX
XX WO9828420-A1.
XX
XX 02-JUL-1998.
XX
XX 18-DEC-1997; 97WO-US23522.
XX
XX 23-DEC-1996; 96US-0034429.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Duan R, Ruben SM;
XX
XX WPI; 1998-377651/32.
XX
XX New nucleic acid encoding human parotid secretory protein or its
XX fragments - useful for diagnosis and treatment of, e.g. digestive
XX and endocrine disorders and for drug screening
XX
XX Disclosure; Page 77; 94pp; English.

```

XX This sequence represents an expressed sequence tag with homology to the
 CC DNA of the invention, and is all so specifically stated as not being
 CC contained within the DNA of the invention. The DNA of the invention
 CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
 CC useful for chromosome identification and isolation of the corresponding
 CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
 CC hPSP (in standard blotting, amplification or immuno assays), particularly
 CC for diagnosis of digestive, non-immune defensive, endocrine or immune
 CC system disorders. A particular application is diagnosis of cancers of the
 CC salivary gland, thymus and pancreas which are associated with high levels
 CC of hPSP. The protein is also useful as antifungal, antibacterial,
 CC antiparasitic and antiviral agents and may be expressed in vivo from the
 CC DNA. The protein, or cells expressing it, are used in screening tests to
 CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
 CC antisense nucleic acids, which are potentially useful for treating
 CC conditions associated with excessive hPSP production. Cells containing
 CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
 CC to identify hPSP-binding proteins.
 XX
 SQ Sequence 374 BP; 86 A; 123 C; 82 G; 78 T; 5 other;

Query Match 20.1%; Score 150; DB 19; Length 374;
 Best Local Similarity 100.0%; Pred. No. 8.8e-66;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 597 GCGTGATCAACACGCTGAAAGCAGCTGTATCTCCCTCTGCAGACGAGATATGTCAC 656
 DB 14 gegtgcacacgcgtgaaaagcactgtatctccctctgctgcagagagatgtccac 73
 QY 657 TGATCCGCATCTTCATCCACTCCCTCGATGTGAATGTCATTGACGAGTCTGCGATAATC 716
 DB 74 tgatccgcattcttccactccctcgatgtgaatgtcattcagcaggtctgataatc 133
 QY 717 CTCAGCACAAACCCAGCTGCAACCCCTCA 746
 DB 134 ctacgacacaaacccagctgcaaacccctca 163

RESULT 22
 AAV44762
 ID AAV44762 standard; DNA; 359 BP.
 XX
 XX AAV44762;
 XX
 DT 16-OCT-1998 (first entry)
 XX
 XX Expressed sequence tag USGSAB9K.
 XX
 DE Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 KW non-immune defensive disorder; immune system disorder; cancer; human;
 KW therapy; diagnosis; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO9828420-A1.
 XX
 PD 02-JUL-1998.
 XX
 XX 18-DEC-1997; 97WO-US23522.
 XX
 XX 23-DEC-1996; 96US-0034429.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Duan R, Ruben SM;
 PI WPI; 1998-377651/32.
 DR
 DR New nucleic acid encoding human parotid secretory protein or its
 PT fragments - useful for diagnosis and treatment of, e.g. digestive

PT and endocrine disorders and for drug screening
 XX
 PS Disclosure; Page 76-77; 94pp; English.
 XX
 CC This sequence represents an expressed sequence tag with homology to the
 CC DNA of the invention, and is all so specifically stated as not being
 CC contained within the DNA of the invention. The DNA of the invention
 CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
 CC useful for chromosome identification and isolation of the corresponding
 CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
 CC hPSP (in standard blotting, amplification or immuno assays), particularly
 CC for diagnosis of digestive, non-immune defensive, endocrine or immune
 CC system disorders. A particular application is diagnosis of cancers of the
 CC salivary gland, thymus and pancreas which are associated with high levels
 CC of hPSP. The protein is also useful as antifungal, antibacterial,
 CC antiparasitic and antiviral agents and may be expressed in vivo from the
 CC DNA. The protein, or cells expressing it, are used in screening tests to
 CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
 CC antisense nucleic acids, which are potentially useful for treating
 CC conditions associated with excessive hPSP production. Cells containing
 CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
 CC to identify hPSP-binding proteins.
 XX
 SQ Sequence 359 BP; 94 A; 79 C; 90 G; 90 T; 6 other;

Query Match 14.7%; Score 110; DB 19; Length 359;
 Best Local Similarity 100.0%; Pred. No. 1.6e-45;
 Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GATGCTTCAGCTTTGGAACCTTGTCTCTCTGTGCGCGCTGTCTACTGGACCTCAGATC 60
 DB 50 gatgcttcagctttggaaactgttctctctgtgcgcggtgctcactggacctcagatc 109
 QY 61 TCTTCTTGCAATCTTGCAATGACCTAAGCAATGCTGTGGATAAGCTGG 110
 DB 110 tcttcttgcaatcttgcaatgacctgaagcaatgctgtggataagctgg 159

RESULT 23
 AAV44769
 ID AAV44769 standard; DNA; 39 BP.
 XX
 XX AAV44769;
 XX
 DT 16-OCT-1998 (first entry)
 XX
 XX PCR primer for human parotid secretory protein coding sequence.
 XX
 DE Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 KW non-immune defensive disorder; immune system disorder; cancer; human;
 KW therapy; diagnosis; PCR primer; ss.
 XX
 XX Synthetic.
 XX
 XX Homo sapiens.
 XX
 XX WO9828420-A1.
 XX
 PD 02-JUL-1998.
 XX
 XX 18-DEC-1997; 97WO-US23522.
 XX
 XX 23-DEC-1996; 96US-0034429.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Duan R, Ruben SM;
 PI WPI; 1998-377651/32.
 DR
 DR New nucleic acid encoding human parotid secretory protein or its
 PT fragments - useful for diagnosis and treatment of, e.g. digestive

PT and endocrine disorders and for drug screening
 XX Example 1; Page 54; 94pp; English.
 PS
 CC This sequence represents a PCR primer for DNA encoding the human parotid
 CC secretory protein (hPSP) of the invention. The hPSP DNA is useful for
 CC chromosome identification and isolation of the corresponding genomic DNA.
 CC The DNA and protein can be used to detect abnormal levels of hPSP (in
 CC standard blotting, amplification or immuno assays), particularly for
 CC diagnosis of digestive, non-immune defensive, endocrine or immune system
 CC disorders. A particular application is diagnosis of cancers of the
 CC salivary gland, thymus and pancreas which are associated with high levels
 CC of hPSP. The protein is also useful as antifungal, antibacterial,
 CC antiparasitic and antiviral agents and may be expressed in vivo from the
 CC DNA. The protein, or cells expressing it, are used in screening tests to
 CC identify specific (antagonists, e.g. antibodies (Ab), polypeptides and
 CC antisense nucleic acids, which are potentially useful for treating
 CC conditions associated with excessive hPSP production. Cells containing
 CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
 CC to identify hPSP-binding proteins.
 XX
 SQ Sequence 39 BP; 9 A; 10 C; 8 G; 12 T; 0 other;
 Query Match 3.8%; Score 28; DB 19; Length 39;
 Best Local Similarity 100.0%; Pred. No. 0.00054;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 56 GAGTCTCTTCTTCACAATCTGGCAATG 83
 DB 12 gagtctctcttgacaatctggcaatg 39
 RESULT 24
 AAV4771
 ID AAV44771 standard; DNA: 43 BP.
 AC AAV44771;
 XX
 DT 16-OCT-1998 (first entry)
 XX
 DE PCR primer for human parotid secretory protein coding sequence.
 XX
 KW Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 KW non-immune defensive disorder; immune system disorder; cancer; human;
 KW therapy; diagnosis; PCR primer; ss.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO9828420-A1.
 XX
 PD 02-JUL-1998.
 XX
 PF 18-DEC-1997; 97WO-US23522.
 XX
 PR 23-DEC-1996; 96US-0034429.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Duan R, Ruben SM;
 XX
 DR WPI; 1998-377651/32.
 XX
 CC New nucleic acid encoding human parotid secretory protein or its
 CC fragments - useful for diagnosis and treatment of, e.g. digestive
 CC and endocrine disorders and for drug screening
 XX
 PS Example 2; Page 58; 94pp; English.
 XX
 CC This sequence represents a PCR primer for DNA encoding the human parotid
 CC secretory protein (hPSP) of the invention. The hPSP DNA is useful for

CC chromosome identification and isolation of the corresponding genomic DNA.
 CC The DNA and protein can be used to detect abnormal levels of hPSP (in
 CC standard blotting, amplification or immuno assays), particularly for
 CC diagnosis of digestive, non-immune defensive, endocrine or immune system
 CC disorders. A particular application is diagnosis of cancers of the
 CC salivary gland, thymus and pancreas which are associated with high levels
 CC of hPSP. The protein is also useful as antifungal, antibacterial,
 CC antiparasitic and antiviral agents and may be expressed in vivo from the
 CC DNA. The protein, or cells expressing it, are used in screening tests to
 CC identify specific (antagonists, e.g. antibodies (Ab), polypeptides and
 CC antisense nucleic acids, which are potentially useful for treating
 CC conditions associated with excessive hPSP production. Cells containing
 CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
 CC to identify hPSP-binding proteins.
 XX
 SQ Sequence 43 BP; 8 A; 13 C; 9 G; 13 T; 0 other;
 Query Match 3.4%; Score 25; DB 19; Length 43;
 Best Local Similarity 100.0%; Pred. No. 0.018;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ATGCTTCAGCTTTGGAACTTGCTC 26
 DB 19 atgcttcagctttggaaacttgctc 43
 RESULT 25
 AAV4766
 ID AAV44766 standard; DNA: 395 BP.
 XX
 AC AAV44766;
 XX
 DT 16-OCT-1998 (first entry)
 XX
 DE Expressed sequence tag HSPMF91R.
 XX
 KW Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 KW non-immune defensive disorder; immune system disorder; cancer; human;
 KW therapy; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9828420-A1.
 XX
 PD 02-JUL-1998.
 XX
 PF 18-DEC-1997; 97WO-US23522.
 XX
 PR 23-DEC-1996; 96US-0034429.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Duan R, Ruben SM;
 XX
 DR WPI; 1998-377651/32.
 XX
 CC New nucleic acid encoding human parotid secretory protein or its
 CC fragments - useful for diagnosis and treatment of, e.g. digestive
 CC and endocrine disorders and for drug screening
 XX
 PS Disclosure; Page 78; 94pp; English.
 XX
 CC This sequence represents an expressed sequence tag with homology to the
 CC DNA of the invention, and is all so specifically stated as not being
 CC contained within the DNA of the invention. The DNA of the invention
 CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
 CC useful for chromosome identification and isolation of the corresponding
 CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
 CC hPSP (in standard blotting, amplification or immuno assays), particularly
 CC for diagnosis of digestive, non-immune defensive, endocrine or immune
 CC system disorders. A particular application is diagnosis of cancers of the

CC salivary gland, thymus and pancreas which are associated with high levels
 CC of hPSP. The protein is also useful as antifungal, antibacterial,
 CC antiparasitic and antiviral agents and may be expressed in vivo from the
 CC DNA. The protein, or cells expressing it, are used in screening tests to
 CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
 CC antisense nucleic acids, which are potentially useful for treating
 CC conditions associated with excessive hPSP production. Cells containing
 CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
 CC to identify hPSP-binding proteins.

XX Sequence 395 BP; 110 A; 116 C; 83 G; 70 T; 16 other;

Query Match 3.2%; Score 24; DB 19; Length 395;
 Best Local Similarity 100.0%; Pred. No. 0.057;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 CATCCACTCCCTGGATGTGAATCT 693
 Db 60 catccactccctggatgtgaatgt 83
 |||||

RESULT 26
 AAV44770/C
 ID AAV44770 standard; DNA; 36 BP.

XX AC AAV44770;

XX DT 16-OCT-1998 (first entry)

XX DE PCR primer for human parotid secretory protein coding sequence.

XX KW Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 XX KW non-immune defensive disorder; immune system disorder; cancer; human;
 XX KW therapy; diagnosis; PCR primer; ss.

XX OS Synthetic.
 XX OS Homo sapiens.

XX PN W09828420-A1.

XX PD 02-JUL-1998.

XX PF 18-DEC-1997; 97WD-US23522.

XX PR 23-DEC-1996; 96US-0034429.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Duan R, Ruben SM;

XX XX WPI: 1998-377651/32.

XX PT New nucleic acid encoding human parotid secretory protein or its
 XX PT fragments - useful for diagnosis and treatment of, e.g. digestive
 XX PT and endocrine disorders and for drug screening

XX PS Example 1; Page 54; 94pp; English.

XX CC This sequence represents a PCR primer for DNA encoding the human parotid
 XX CC secretory protein (hPSP) of the invention. The hPSP DNA is useful for
 XX CC chromosome identification and isolation of the corresponding genomic DNA.
 XX CC The DNA and protein can be used to detect abnormal levels of hPSP (in
 XX CC standard blotting, amplification or immuno assays), particularly for
 XX CC diagnosis of digestive, non-immune or immune disorders.
 XX CC disorders. A particular application is diagnosis of cancers of the
 XX CC salivary gland, thymus and pancreas which are associated with high levels
 XX CC of hPSP. The protein is also useful as antifungal, antibacterial,
 XX CC antiparasitic and antiviral agents and may be expressed in vivo from the
 XX CC DNA. The protein, or cells expressing it, are used in screening tests to
 XX CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
 XX CC antisense nucleic acids, which are potentially useful for treating

CC conditions associated with excessive hPSP production. Cells containing
 CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and
 CC to identify hPSP-binding proteins.

XX Sequence 36 BP; 6 A; 7 C; 12 G; 11 T; 0 other;

Query Match 2.9%; Score 22; DB 19; Length 36;
 Best Local Similarity 100.0%; Pred. No. 0.59;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 725 AAAACCCAGCTGCAACCCCTCA 746
 Db 36 AAAACCCAGCTGCAACCCCTCA 15
 |||||

RESULT 27

AAH18450/C

ID AAH18450 standard; cDNA; 1730 BP.

XX AC AAH18450;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:18545.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX XX WPI: 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602

XX PT full-length cDNAs defined in the specification, and for the detection

XX PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs -

XX PS Claim 8; SEQ ID 18545; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602
 XX CC full-length cDNAs defined in the specification. Where a primer set
 XX CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 XX CC to the complementary strand of a polynucleotide which comprises one of
 XX CC the 5602 nucleotide sequences defined in the specification, where the
 XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 XX CC of an oligonucleotide comprising a sequence complementary to the
 XX CC complementary strand of a polynucleotide which comprises a 5'-end
 XX CC sequence and an oligonucleotide comprising a sequence complementary to a
 XX CC polynucleotide which comprises a 3'-end sequence, where the
 XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
 XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
 XX CC the specification. The primer sets can be used in antisense therapy and
 XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 XX CC particularly full-length cDNAs. The primers are also useful for the
 XX CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH16742 represent human cDNA sequences; AA892446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 1730 BP; 547 A; 358 C; 383 G; 442 T; 0 other;

Query Match 2.8%; Score 21; DB 22; Length 1730;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 TGAGAAATTCCTGAACAATGT 264
 |||||
 DB 1113 TGAGAAATTCCTGAACAATGT 1093

RESULT 28
 AAV44768
 ID AAV44768 standard; DNA; 360 BP.

XX
 AC AAV44768;

DT 16-OCT-1998 (first entry)

XX Expressed sequence tag HSPMF57R.

XX Parotid secretory protein; hPSP; digestive disorder; endocrine disorder;
 KW non-immune defensive disorder; immune system disorder; cancer; human;
 KW therapy; diagnosis; ss.

XX Homo sapiens.

XX WO9828420-A1.

XX 02-JUL-1998.

XX 18-DEC-1997; 97WO-US23522.

XX 23-DEC-1996; 96US-0034429.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Duan R, Ruben SM;

XX WPI; 1998-377651/32.

XX New nucleic acid encoding human parotid secretory protein or its
 PT fragments - useful for diagnosis and treatment of, e.g. digestive
 PT and endocrine disorders and for drug screening

XX Disclosure: Page 79; 94pp; English.

XX This sequence represents an expressed sequence tag with homology to the
 CC DNA of the invention, and is all so specifically stated as not being
 CC contained within the DNA of the invention. The DNA of the invention
 CC encodes the human parotid secretory protein (hPSP). The hPSP DNA is
 CC useful for chromosome identification and isolation of the corresponding
 CC genomic DNA. The DNA and protein can be used to detect abnormal levels of
 CC hPSP (in standard blotting, amplification or immuno assays), particularly
 CC for diagnosis of digestive, non-immune defensive, endocrine or immune
 CC system disorders. A particular application is diagnosis of cancers of the
 CC salivary gland, thymus and pancreas which are associated with high levels
 CC of hPSP. The protein is also useful as antifungal, antibacterial,
 CC antiparasitic and antiviral agents and may be expressed in vivo from the
 CC DNA. The protein, or cells expressing it, are used in screening tests to
 CC identify specific (ant)agonists, e.g. antibodies (Ab), polypeptides and
 CC antisense nucleic acids, which are potentially useful for treating
 CC conditions associated with excessive hPSP production. Cells containing
 CC the DNA are used to express the recombinant protein and this can be used
 CC to raise Ab, useful for diagnosis, therapy, for affinity purification and

CC to identify hPSP-binding proteins.

XX
 SQ Sequence 360 BP; 99 A; 91 C; 83 G; 60 T; 27 other;

Query Match 2.7%; Score 20; DB 19; Length 360;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 TTATCAACCTGAAAGCTCC 448
 |||||
 DB 30 ttatcaacctgaaagcctcc 49

RESULT 29
 AAV04633
 ID AAV04633 standard; cDNA; 4692 BP.

XX
 AC AAV04633;

DT 17-AUG-1998 (first entry)

XX Porcine phosphoinositide 3OH-kinase p101 subunit cDNA.

XX G-beta-gamma regulated phosphatidylinositol-3' kinase; pig;
 KW phosphoinositide 3OH-kinase; PI3K; signal transduction;
 KW phosphatidylinositol (3,4,5)-triphosphate; G-protein; receptor;
 KW transgenic animal; knockout animal; inflammation; arthritis;
 KW septic shock; adult respiratory distress syndrome; pneumonia;
 KW asthma; allergy; reperfusion injury; atherosclerosis; cancer;
 KW Alzheimer's disease; cancer; antisense; ribozyme; diagnosis;
 KW therapy; drug screening; ss.

XX Sus scrofa.

XX Key Location/Qualifiers
 FH 390..3023
 FT CDS /*tag= a

XX WO9749818-A2.

XX 31-DEC-1997.

XX 26-JUN-1997; 97WO-US11219.

XX 27-JUN-1996; 96US-0672211.

XX (ONYX-) ONYX PHARM.

XX Braselmann S, Hawkins PT, Stephens L;

XX WPI; 1998-077181/07.

XX P-PSDB; AAW23946.

XX DNA encoding G-beta-gamma regulated phosphatidylinositol-3' kinase,
 PT p101 and p120 subunits - useful for diagnosis, drug screening,
 PT clinical trial monitoring and treatment of inflammatory disorders
 XX Claim 2; Fig 1A-D; 151pp; English.

XX This cDNA sequence codes for the p101 regulatory subunit (see
 CC AAW23946) of pig G-protein regulated phosphatidylinositol-3' kinase,
 CC a heterodimeric enzyme which produces the intracellular messenger
 CC phosphatidylinositol (3,4,5)-triphosphate in response to activation
 CC of trimeric G protein-linked receptors. This novel protein, which
 CC also contains a catalytic subunit, p120 (see AAV04634), is found in
 CC cells of haematopoietic origin and is involved in immune system
 CC responses which cause inflammation. The p101 subunit is largely
 CC responsible for the dramatic stimulation of kinase activity in the
 CC presence of activated trimeric C proteins. p101 cDNA was obtained
 CC from a pig neutrophil cDNA library using a degenerate oligonucleotide
 CC probe (see AAW99714) based on an isolated p101 tryptic peptide. The
 CC p101 cDNA clone in pCMV3mycpl01 is deposited as ATCC 97636. The

CC invention encompasses pig and human p101 and p120 nucleotides,
 CC host cell expression systems, p101 and p120 proteins (see
 CC AAM23946-49), fusion proteins, polypeptides and peptides, antibodies
 CC to these proteins, and transgenic animals and knockout animals.
 CC Compounds which are useful for treating inflammatory response
 CC disorders can be identified by screening assays using a G protein
 CC activated P13K, or a cultured host cell that expresses the p101
 CC gene. Antagonists of G protein stimulated P13K (acting through the
 CC p101 subunit, especially by disrupting the interaction between the
 CC p101 and p120 subunits) can be used to treat arthritis, septic
 CC shock, adult respiratory distress syndrome (ARDS), pneumonia,
 CC asthma, allergies, reperfusion injury, atherosclerosis, cancer and
 CC Alzheimer's disease. Signalling-incompetent p101 (preferably a
 CC deletion mutant) oligonucleotide constructs, antisense
 CC oligonucleotides, triple helix forming oligonucleotides, ribozymes
 CC or recombinant DNA constructs that act through targeted homologous
 CC recombination are all useful for inhibiting expression of the p101
 CC subunit, especially by delivery to cells of haematopoietic origin.
 CC (All claimed). The nucleic acids and their products can also be
 CC used for diagnosis, drug screening and clinical trial monitoring of
 CC inflammatory diseases.

XX
 SQ Sequence 4692 BP; 916 A; 1559 C; 1394 G; 823 T; 0 other;

Query Match 2.7%; Score 20; DB 19; Length 4692;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 665 ATCTTCATCCACTCCCTCGA 684
 |||||||
 Db 2448 atcttcacccactccctgga 2467

RESULT 30
 AAV74099
 ID AAV74099 standard; DNA; 4692 BP.
 AC
 XX
 XX
 DT 09-MAR-1999 (first entry)
 DE Porcine G-protein regulated P13K p101 adapter subunit DNA.
 XX Phosphoinositide 3-hydroxylase; P13K; trimeric G protein; porcine;
 KW adapter subunit; regulatory subunit; p101; p120; catalytic subunit;
 KW detection; diagnosis; activation disorder; haematopoietic system;
 KW treatment; immune disorder; inflammation; arthritis; septic shock;
 KW adult respiratory distress syndrome; pneumonia; asthma; allergy;
 KW reperfusion injury; atherosclerosis; Alzheimer's disease; cancer; ss.

XX
 OS Sus scrofa.
 XX
 PN US5856132-A.
 XX
 PD 05-JAN-1999.
 XX
 PF 15-AUG-1997; 97US-0916917.
 XX
 PR 15-AUG-1997; 97US-0916917.
 XX
 PR 27-JUN-1996; 96US-0672211.
 XX
 PA (ONYX-) ONYX PHARM.
 XX
 PI Braselmann S, Hawkins PT, Stephens L;
 XX
 DR WPI: 1999-105107/09.
 XX
 DR P-PSDB; AAW90082.
 XX
 XX Nucleic acid encoding regulatory (p101) and catalytic (p120)
 PT subunits of a heterodimeric phosphatidylinositol-3' kinase - useful
 PT in treatment and diagnosis of immune system disorders, e.g.
 PT arthritis, cancer and Alzheimer's disease

XX
 PS Example IX; Fig 1A-D; 75pp; English.
 CC
 CC This sequence encodes a novel adapter subunit, p101, (also known as the
 CC regulatory subunit) from porcine phosphoinositide 3-hydroxylase
 CC (P13K) which is regulated by beta-gamma subunits of trimeric G-protein.
 CC Nucleic acid coding for p101 or p120, or their fragments, are used
 CC as probes and primers for identifying p101 or p120 gene mutations,
 CC allelic variations or regulatory defects, particularly for the diagnosis
 CC of activation disorders (or susceptibility) in cells of the
 CC haematopoietic system. The related proteins, antibodies, agonists and
 CC antagonists can be used similarly. The p101 and p120 proteins, peptides
 CC or fusion proteins are used to treat or screen for potential agents for
 CC treating immune disorders, particularly inflammation, e.g. arthritis,
 CC septic shock, adult respiratory distress syndrome, pneumonia, asthma,
 CC allergies, reperfusion injury, atherosclerosis, Alzheimer's disease and
 CC cancer.

XX
 SQ Sequence 4692 BP; 916 A; 1559 C; 1394 G; 823 T; 0 other;

Query Match 2.7%; Score 20; DB 20; Length 4692;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 665 ATCTTCATCCACTCCCTCGA 684
 |||||||
 Db 2448 atcttcacccactccctgga 2467

RESULT 31
 AA286812
 ID AA286812 standard; DNA; 4692 BP.
 XX
 AC AA286812;
 XX
 DT 20-APR-2000 (first entry)
 XX
 DE Pig p101 regulatory subunit coding sequence.
 XX
 KW p101 protein; p120 protein; regulatory subunit; immune system disorder;
 KW trimeric G-protein regulated P13K; phosphoinositide 3OH-kinase; asthma;
 KW inflammatory response disorder; arthritis; septic shock; allergy;
 KW adult respiratory distress; cancer; reperfusion injury; atherosclerosis;
 KW Alzheimer's disease; haematopoietic lineage cell activation disorder;
 KW therapy; pig; ss.

XX
 OS Sus sp.
 XX
 PN US6017763-A.
 XX
 PD 25-JAN-2000.
 XX
 PF 04-JAN-1999; 99US-0225170.
 XX
 PR 15-AUG-1997; 97US-0916917.
 XX
 PR 27-JUN-1996; 96US-0672211.
 XX
 PA (ONYX-) ONYX PHARM INC.
 XX
 PA (BABR-) BABRAMHAM INST.
 XX
 PI Braselmann S, Stephens L, Hawkins PT;
 XX
 DR WPI: 2000-136682/12.
 XX
 XX Isolated p101 regulatory polypeptide, a subunit of the trimeric
 PT G-protein-regulated P13K, is useful for screening compounds which can
 PT be used to treat inflammatory response disorders -
 XX
 XX Example; Fig 1; 75pp; English.
 PS
 XX This sequence encodes the pig p101 regulatory protein.
 CC
 CC The invention relates to the human p101 regulatory protein,

PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SC1 INF.

Rosen CA, Harash SC, Ruben SM;

WPI; 2001-541565/60.

P-PSDB; ABB14727.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Claim 1; SEQ ID NO 60; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABA11004-ABA21534) and proteins (AB14678-AB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 425 BP; 102 A; 111 C; 80 G; 120 T; 12 other;

Query Match 2.5%; Score 19; DB 22; Length 425;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 367 CCTAACCTGAGCTCCCT 385
|||||
DB 210 ccttaacctgagctccct 228

RESULT 33

ABAL4447
ID ABA14447 standard; DNA; 451 BP.

XX
AC ABA14447;

XX
DT 23-JAN-2002 (first entry)

XX
DE Human nervous system related polynucleotide SEQ ID NO 6778.

XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX
OS Homo sapiens.

XX
PN WO200159063-A2.

XX
PD 16-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US01334.

XX
PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.


```
Db 210 ccttaacctgagcttcccl 228
|||||
RESULT 34
ABA14448
ID ABA14448 standard: DNA; 766 BP.
XX
XX
AC ABA14448;
XX
XX
DT 23-JAN-2002 (first entry)
XX
XX Human nervous system related polynucleotide SEQ ID NO 6779.
XX
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antitubercular; vulnerary;
KW antiparkinsonian; antisticking; antianaemic; antithrombotic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antifungal; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; neurotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
```

PR 17-NOV-2000: 2000US-0249244.
 PR 17-NOV-2000: 2000US-0249245.
 PR 17-NOV-2000: 2000US-0249246.
 PR 17-NOV-2000: 2000US-0249265.
 PR 17-NOV-2000: 2000US-0249297.
 PR 17-NOV-2000: 2000US-0249299.
 PR 17-NOV-2000: 2000US-0249300.
 PR 01-DEC-2000: 2000US-0250391.
 PR 01-DEC-2000: 2000US-0251160.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.
 PR 05-DEC-2000: 2000US-0256719.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251989.
 PR 08-DEC-2000: 2000US-0251990.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI: 2001-541565/60.
 XX
 XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases.
 XX
 XX Disclosure: SEQ ID NO 6779; 1701pp + Sequence Listing; English.
 XX
 XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins
 CC (ABAI4678-ABAI6001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 766 BP; 165 A; 221 C; 155 G; 225 T; 0 other;
 SQ
 Query Match 2.5%; Score 19; DB 22; Length 766;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 367 CCTTAACCTGAGCTTCCTC 385
 DB 525 ccttaacctgagcttcctc 543
 RESULT 35
 AAV18462/C
 ID AAV18462 standard; DNA: 2957 BP.
 XX
 XX AAV18462;
 AC
 XX
 XX 01-OCT-1998 (first entry)
 DT
 XX Human granulocytic ehrlichiosis isolate NCH-1 eM4 DNA sequence.
 DE

XX
 KW Human: granulocytic ehrlichiosis; HGE; aOHGE; diagnosis; vaccine;
 KW detection; infection; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 1330..2325
 FT /*tag= a
 FT /product= "eM4 protein"
 XX
 XX WO9814584-A2.
 PN
 XX 09-APR-1998.
 PD
 XX 30-SEP-1997; 97WO-US17675.
 PF
 XX 01-OCT-1996; 96US-0027180.
 PR
 XX (UYA) UNIV YALE.
 PA
 XX Barthold SW, Fikrig E, Ijdo J, Sun W;
 PI
 XX WPI: 1998-260999/23.
 DR P-PSDB; AAW48748.
 DR
 XX New isolated human granulocyte ehrlichiosis DNA and proteins - used
 PT to develop products for detection, treatment and prevention of human
 PT granulocytic ehrlichiosis and related disorders
 XX
 XX Example 8; Fig 11; 159pp; English.
 PS
 XX The present sequence encodes the human granulocytic ehrlichiosis
 CC (HGE) isolate NCH-1 eM4 protein. The present invention describes
 CC HGE DNA molecules and proteins, and methods which can be used for the
 CC detection, treatment and prevention of HGE and related disorders caused
 CC by infection by protein and DNA molecules which encode them, from the
 CC agent of HGE (aOHGE). An antibody which binds to an HGE protein can
 CC be used in vaccines.
 XX
 XX Sequence 2957 BP; 874 A; 506 C; 829 G; 747 T; 1 other;
 SQ
 Query Match 2.5%; Score 19; DB 19; Length 2957;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 310 AATCAGCAACTCCCTCATC 328
 DB 991 AATCAGCAACTCCCTCATC 973
 RESULT 36
 AAS91657/C
 ID AAS91657 standard; CDNA: 3636 BP.
 XX
 AC AAS91657;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 XX DNA encoding novel human diagnostic protein #27461.
 DE
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX

```

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WP1: 2001-639362/73.
DR P-PSDB: ABG27470.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1: SEQ ID NO 27461; 103pp: English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3636 BP; 585 A; 1224 C; 1162 G; 665 T; 0 other;
XX
XX
XX Query Match 2.5%; Score 19; DB 23; Length 3636;
XX Best: Local Similarity 100.0%; Pred. No. 20;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 227 CAGAGGCCCGCAGAGCTG 245
Db 1539 CAGAGGCCCGCAGAGCTG 1521
IIIIIIIIIIIIIIIIIIII
RESULT 37
AAK85994
ID AAK85994 standard; DNA; 59060 BP.
XX
XX AAK85994;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40806.
DE
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytototoxic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
PN
XX 09-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01354.
PF
XX 31-JAN-2000; 2000US-0179065.
PR

```

```

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

```


CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 846 BP; 196 A; 215 C; 241 G; 194 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 846;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 ACCGCGATCACAATGAA 478

Db 26 ACCGCGATCACAATGAA 9

RESULT 39

AA553405
 ID AA553405 standard; DNA; 1173 BP.

AC AA553405;

DT 13-FEB-2002 (first entry)

DE Haemophilus influenzae DNA for cellular proliferation protein #187.

KW Antisense; ds: prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.

OS Haemophilus influenzae.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

XX P-PSDB: AAU35546.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 7042; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 1173 BP; 362 A; 205 C; 281 G; 325 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 1173;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 GAGAAATTGCTGACAAAT 262

Db 893 gagaaattgctgacaaat 910

RESULT 40

AAH73917

ID AAH73917 standard; cDNA; 1274 BP.

AC AAH73917;

DT 04-OCT-2001 (first entry)

DE Human PDZ protein 20 coding sequence.

KW Human; PDZ protein 20; cytostatic; virucidal; immunomodulatory;
 KW antiinflammatory; haemostatic; malignant tumour; haemopathy;
 KW HIV infection; immunological disease; inflammation; ss.

OS Homo sapiens.

PN WO200155416-A1.

PD 02-AUG-2001.

PF 15-JAN-2001; 2001WO-CN00036.

PR 26-JAN-2000; 2000CN-0111536.

PA (BIOD-) BIODOR GENE TECHNOLOGY LTD SHANGHAI.

PI Mao Y, Xie Y;

XX WPI: 2001-476218/51.

XX P-PSDB: AAG64378.

XX Human PDZ protein 20 and encoded polynucleotide, applicable in
 PT diagnosis and treatment of malignant tumor, hemopathy, HIV infection,
 PT immunological diseases and various inflammations -

PS Claim 6; Page 29-30; 34pp; Chinese.

XX The present sequence is the coding sequence for human PDZ protein 20.
 CC PDZ protein 20 and its coding sequence are useful in the diagnosis and
 CC treatment of malignant tumour, haemopathy, HIV infection, immunological
 CC diseases and various inflammations.

SQ Sequence 1274 BP; 301 A; 394 C; 364 G; 215 T; 0 other;

Query Match 2.4%; Score 18; DB 22; Length 1274;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 CAGCAACTCCCTCATCCT 330

Db 780 cagcaactccctcatcct 797

RESULT 41
 AAS46228/c
 ID AAS46228 standard; cDNA: 2265 BP.

XX AC AAS46228;

XX DT 18-DEC-2001 (first entry)

XX DE Human DNA encoding PRO polypeptide sequence #304.

XX KW PRO polypeptide: mammal; tumour; cancer; human; cattle; horse; sheep; ss;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
 KW PCR primer.

XX OS Homo sapiens.

XX PN WO200158848-A2.

XX PD 20-SEP-2001.

XX PF 28-FEB-2001; 2001WO-US06520.

XX PR 01-MAR-2000; 2000WO-US05601.

XX PR 02-MAR-2000; 2000WO-US05841.

XX PR 03-MAR-2000; 2000US-187202P.

XX PR 06-MAR-2000; 2000US-186968P.

XX PR 14-MAR-2000; 2000US-189320P.

XX PR 14-MAR-2000; 2000US-189328P.

XX PR 15-MAR-2000; 2000WO-US06884.

XX PR 21-MAR-2000; 2000US-190828P.

XX PR 21-MAR-2000; 2000US-191007P.

XX PR 21-MAR-2000; 2000US-191048P.

XX PR 21-MAR-2000; 2000US-191314P.

XX PR 28-MAR-2000; 2000US-192855P.

XX PR 29-MAR-2000; 2000US-193032P.

XX PR 29-MAR-2000; 2000US-193053P.

XX PR 30-MAR-2000; 2000WO-US08439.

XX PR 04-APR-2000; 2000US-194449P.

XX PR 04-APR-2000; 2000US-194647P.

XX PR 11-APR-2000; 2000US-195975P.

XX PR 11-APR-2000; 2000US-196000P.

XX PR 11-APR-2000; 2000US-196187P.

XX PR 11-APR-2000; 2000US-196690P.

XX PR 18-APR-2000; 2000US-196820P.

XX PR 18-APR-2000; 2000US-198121P.

XX PR 25-APR-2000; 2000US-198585P.

XX PR 25-APR-2000; 2000US-198397P.

XX PR 25-APR-2000; 2000US-199550P.

XX PR 25-APR-2000; 2000US-199654P.

XX PR 03-MAY-2000; 2000US-201516P.

XX PR 17-MAY-2000; 2000WO-US13705.

PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -

XX Claim 2; Fig 607; 774pp; English.

XX Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
 CC primers for PRO polypeptides of the invention. The sequences of the
 CC invention can be used to detect the presence of a tumour in a mammal by
 CC comparing the level of expression of a PRO polypeptide in a test sample
 CC of cells from the animal and a control sample of normal cells, whereby a
 CC higher level of expression in the test sample indicates the presence of a
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human
 CC blood, when contacted with it. A specific polypeptide can be used to
 CC stimulate the proliferation or differentiation of chondrocyte cells. The
 CC PRO proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.

XX Sequence 2265 BP; 542 A; 635 C; 601 G; 487 T; 0 other;

Query Match 2.4%; Score 18; DB 22; Length 2265;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 GGGCCATCATGGCCAGA 428

|||||

Db 1908 GGGCCATCATGGCCAGA 1891

RESULT 42

AAAX18951/c

ID AAAX18951 standard; cDNA: 2345 BP.

XX AC AAAX18951;

XX DT 13-MAY-1999 (first entry)

XX DE Human PIGR-1 encoding cDNA.

XX Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;

XX multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;

XX inflammatory bowel disease; ss.

XX OS Homo sapiens.

XX PN EP897981-A1.

XX PD 24-FEB-1999.

XX PF 11-AUG-1998; 98EP-0306403.

XX PR 22-OCT-1997; 97US-0955937.

XX PR 19-AUG-1997; 97US-0056152.

XX PA (SMIK) SMITHKLINE BEECHAM.

XX PI Hurle MR, Sweet RW, Truneh A, Wu S;

XX WPI: 1999-134644/12.

XX P-PSDB; AA999070.

XX PT New receptor (PIGR-1) polypeptide and polynucleotide - useful as

XX diagnostic reagents and for prevention and treatment of multiple

XX sclerosis, inflammatory bowel disease and psoriasis

XX Claim 4; Page 6; 28pp; English.

(GETH) GENENTECH INC.

Baker KP, Chen J, Goddard A, Godowsky PJ, Gurney AL;

Pan J, Smith V, Watanabe CK, Wood WI, Zhang 2;

WPI: 2001-602746/68.

P-PSDB; AAU29327.

The present sequence encodes a new receptor polypeptide designated PIGR-1, which is a member of the immunoglobulin (Ig) superfamily. PIGR-1 proteins and polynucleotides are useful for diagnosing susceptibility to diseases by detecting mutations in the PIGR-1 gene, and can diagnose diseases associated with PIGR-1 protein imbalance by determining PIGR-1 protein expression levels. PIGR-1 proteins can be used to screen for agonists and antagonists by measuring the binding to protein, and observing the protein function. These can be used in treatment to activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition to direct administration of antisense sequences to prevent expression, or PIGR-1 polynucleotides to treat conditions associated with a lack of PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1 protein expression. PIGR-1 antibodies are useful for inducing an immune response to immunise and prevent disease, and for isolating PIGR-1 clones or purifying the polypeptides by affinity chromatography. PIGR-1 proteins can be administered directly or as a vaccine to inoculate against disease. Diseases diagnosed, prevented and treated include: rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus erythematosus and inflammatory bowel disease. The PIGR-1 protein is also useful for mapping the gene to a chromosome, allowing gene inheritance to be studied through linkage analysis.

Sequence 2345 BP: 562 A: 655 C: 628 G: 500 T: 0 other;

Query Match 2.4%; Score 18; DB 20; Length 2345;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 GCCCATCATTTGGCCAGCAGA 428
|||||

DB 1986 GCCCATCATTTGGCCAGCAGA 1969

RESULT 43

ABL17740/C
ID ABL17740 standard; DNA: 2643 BP.

XX ABL17740;
AC

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4693.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene: ds.

XX Drosophila melanogaster.

OS WO200171042-A2.

PX 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Claim 1; SEQ ID NO 4693; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB557737-AB572072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 2643 BP: 756 A: 531 C: 561 G: 795 T: 0 other;

Query Match 2.4%; Score 18; DB 23; Length 2643;
Best Local Similarity 100.0%; Pred. No. 63;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 ACCGCGATCACAATTGAA 478
|||||

DB 2598 ACCGCGATCACAATTGAA 2581

RESULT 44

AAT91872/C
ID AAT91872 standard; cDNA to mRNA; 3224 BP.

XX AAT91872;

AC AAT91872;

DT 28-JAN-1998 (first entry)

XX Rat YT521 gene.

XX rat; YT521 gene; re-oxygenation; astrocyte; low-oxygen condition; brain;

KW ischaemia; animal model; drug screening; ds.

XX Rattus rattus.

XX Key Location/Qualifiers

FT CDS 317..2455

FT /*tag= a

XX JP09238685-A.

PD 16-SEP-1997.

XX 07-MAR-1996; 96JP-0049380.

XX 07-MAR-1996; 96JP-0049380.

PA (TANA) TANABE SEIYAKU CO.

PA (TOYAMA) TOYAMA W.

XX WPI; 1997-506554/47.

DR P-PSDB; AAW30749.

XX Brain ischaemia-related gene and protein - useful for elucidation of
XX mechanisms of ischaemia and in animal models for screening drugs for
XX prevention or treatment of ischaemia

PS Claim 3; Pages 8-9; 14pp; Japanese.

XX The present sequence is the rat YT521 gene which encodes a protein which
XX is expressed specifically upon re-oxygenation of astrocytes after
XX exposure to low-oxygen conditions. The protein is involved in brain
XX ischaemia, so it and the related gene are useful for elucidation of the
XX mechanism of ischaemia. They are also useful for the creation of a model
XX animal with diseases accompanying ischaemia or for development of drugs
XX for prevention or treatment of such diseases. Antibodies to the protein
XX are useful for determining the severity of diseases caused by ischaemia.

XX Sequence 3224 BP: 996 A: 609 C: 823 G: 796 T: 0 other;

Query Match

2.4%; Score 18; DB 18; Length 3224;


```

Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Gaps 0;

Qy 660 TCGGCATCTTCATCCACT 677
    |||||
Db 1002 TCGGCATCTTCATCCACT 985

RESULT 45
ABL09202/c
ID ABL09202 standard; cDNA: 3839 BP.
XX
AC ABL09202;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 22088.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene: ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
PI WPI: 2001-658860/75.
DR
DR P-PSDB; AB863099.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
PT Claim 1: SEQ ID NO 22088; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB857737-AB872072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3839 BP: 1086 A: 878 C: 827 G: 1048 T: 0 other;
SQ

Query Match 2.4%; Score 18; DB 23; Length 3839;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Gaps 0;

Qy 412 GCCCATCATTCGCCAGAT 429
    |||||
Db 3264 GCCCATCATTCGCCAGAT 3247

RESULT 46
ABL18573
ID ABL18573 standard; DNA: 4331 BP.
XX

```

PD 27-SEP-2001.
 XX
 PE 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI: 2001-655860/75.
 DR P-PSDB; ABB68048.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 30935; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU16177), expressed DNA
 CC sequences (ABU161840-ABU16175) and the encoded proteins
 CC (ABU57737-ABU72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 4360 BP; 1145 A; 1067 C; 1110 G; 1038 T; 0 other;
 SQ

 Query Match 2.4%; Score 18; DB 23; Length 4360;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 412 GCCATCATTCGCCAGAT 429
 Db 4009 gccatcatctggccagat 4026
 ||||||||||||||||

 RESULT 48
 AAH18649
 ID AAH18649 standard; cDNA; 4372 BP.
 XX
 XX AAH18649;
 AC
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:18880.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EF1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 XX
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0118767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI: 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 XX Claim 8; SEQ ID 18880; 2537pp + CD ROM; English.
 PS
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide comprising one of
 CC the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 4372 BP; 1270 A; 862 C; 1032 G; 1208 T; 0 other;
 SQ

 Query Match 2.4%; Score 18; DB 22; Length 4372;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 233 GCCCAGGAGACCTGAGAAA 250
 Db 2893 gccccaggaagctgagaaa 2910
 ||||||||||||||||

 RESULT 49
 ABL17748/5
 ID ABL17748 standard; DNA; 5098 BP.
 XX
 XX ABL17748;
 AC
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 4587.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI

XX WPI: 2001-656860/75.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1: SEQ ID NO 4687; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 5098 BP; 1417 A; 1064 C; 1100 G; 1517 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 5098;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 ACCGCAGTCACAAATTGAA 478
 |||||
 Db 1026 ACCGCAGTCACAAATTGAA 1009

RESULT 50
 ABL08533
 ID ABL08533 standard; cDNA: 6642 BP.
 XX
 AC ABL08533;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20081.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WC200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI: 2001-656860/75.
 XX P-PSDB; ABB64430.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Claim 1: SEQ ID NO 20081; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6642 BP; 1919 A; 1663 C; 1673 G; 1387 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 6642;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 533 CCACCAGCATCTCACTT 550
 |||||
 Db 581 ccaaccagcatctcactt 698

RESULT 51
 ABL12150/c
 ID ABL12150 standard; cDNA: 6846 BP.
 XX
 AC ABL12150;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 30932.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WC200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI: 2001-656860/75.
 XX P-PSDB; ABB68047.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Claim 1: SEQ ID NO 30932; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6846 BP; 1790 A; 1630 C; 1592 G; 1834 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 6846;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 412 GCCCATCATGGCCAGAT 429
Db 1352 GCCCATCATGGCCAGAT 1335
|||||
RESULT 52
ABLI18572/C
ID ABLI18572 standard; DNA: 6872 BP.
XX
AC ABLI18572;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7189.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
FA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1: SEQ ID NO 7189; 21pp + Sequence Listing: English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI1840-ABLI16175) and the encoded proteins
CC (ABBI5737-ABBI72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6872 BP: 1795 A; 1639 C; 1598 G; 1840 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 6872;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 412 GCCCATCATGGCCAGAT 429
Db 1352 GCCCATCATGGCCAGAT 1335
|||||
RESULT 53
AAS32860/C
ID AAS32860 standard; DNA: 8115 BP.
XX

AC AAS32860;
XX 17-DEC-2001 (first entry)
DT
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID NO 814.
XX
KW Human; endocrine antigen; ds; cytostatic; antifertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01335.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.

```
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0261476.
PR 08-DEC-2000; 2000US-0261856.
PR 08-DEC-2000; 2000US-0261868.
PR 08-DEC-2000; 2000US-0261869.
PR 08-DEC-2000; 2000US-0261989.
PR 08-DEC-2000; 2000US-0261990.
PR 11-DEC-2000; 2000US-0264097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457726/49.
XX
XX Isolated polypeptide for treating, preventing and prognosing disorders
XX related to the endocrine system including endocrine disorders,
XX reproductive disorders, and gastrointestinal disorders and also for
XX testing and detection e.g. diagnosis -
XX
XX Disclosure; SEQ ID No 814; 558pp; English.
XX
XX The invention relates to cDNAs encoding novel human endocrine
XX antigens or a fragment having biological activity, a domain, an epitope,
XX full length protein, variant, allelic variant or a species homologue of
XX the cDNA/antigen. The DNAs and polypeptides are useful for preventing,
XX treating or ameliorating a medical condition when administered
XX (e.g. by gene therapy or antisense-therapy). Identifying mutations in
XX the genes coding for the antigens is useful for diagnosing a pathological
XX condition or a susceptibility to a pathological condition. The DNAs,
XX antigens and antibodies raised against the antigens useful for treating,
XX preventing and/or prognosing disorders related to the endocrine system
XX or hormone imbalance or reproductive disorders, cancers of endocrine
XX tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal
XX glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
XX hypothalamus and testes (e.g. vanishing testes syndrome), many examples
XX of diseases and disorders are given in the specification. The present
XX sequence is genomic DNA fragment form a gene encoding an endocrine
XX antigen of the invention.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 815 BP; 2052 A; 1961 C; 1841 G; 2261 T; 0 other;
XX

Query Match 2.4%; Score 18; DB 22; Length 8115;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 164 CTTGAGAACTGAGGTC 181
DB 612 CTTGAGAACTGAGGTC 595
|||||
RESULT 54
AAS32861/C
ID AAS32861 standard; DNA; 8116 BP.
XX
XX AAS32861;
XX
XX 17-DEC-2001 (first entry)
XX
XX Human genomic DNA for novel endocrine antigen, SEQ ID No 815.
XX
XX Human; endocrine antigen; ds; cytostatic; antifertility; antidiabetic;
XX thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
XX antisense-therapy; antibody; endocrine disorder; hormone imbalance;
XX reproductive disorder; endocrine cancer; pancreatic disorder;
```

KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX Homo sapiens.
XX WO200155319-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01335.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218390.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239535.
PR 13-OCT-2000; 2000US-0239537.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254037.

PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX WPI: 2001-457726/49.
 XX Isolated polypeptide for treating, preventing and prognosing disorders
 PT related to the endocrine system including endocrine disorders,
 PT reproductive disorders, and gastrointestinal disorders and also for
 PT testing and detection e.g. diagnosis.
 XX Disclosure: SEQ ID NO 815; 558pp; English.
 PS
 XX The invention relates to cDNAs encoding novel human endocrine
 CC antigens or a fragment having biological activity, a domain, an epitope,
 CC full length protein, variant, allelic variant or a species homologue of
 CC the cDNA/antigen. The DNAs and polypeptides are useful for preventing,
 CC treating or ameliorating a medical condition when administered
 CC (e.g. by gene therapy or antisense-therapy). Identifying mutations in
 CC the genes coding for the antigens is useful for diagnosing a pathological
 CC condition or a susceptibility to a pathological condition. The DNAs,
 CC antigens and antibodies raised against the antigens useful for treating,
 CC preventing and/or prognosing disorders related to the endocrine system
 CC or hormone imbalance or reproductive disorders, cancers of endocrine
 CC tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal
 CC glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
 CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples
 CC of diseases and disorders are given in the specification. The present
 CC sequence is genomic DNA fragment form a gene encoding an endocrine
 CC antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 8116 BP; 2054 A; 1957 C; 1843 G; 2262 T; 0 other;

Query Match 2.4%; Score 18; DB 22; Length 8116;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 164 CTTGAGAACTCAAGGTC 181
 DB 612 CTTGAGAACTCAAGGTC 595
 |||||
 RESULT 55
 ABL08532/C
 ID ABL08532 standard; CDNA; 8844 BP.
 XX
 AC ABL08532;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20078.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11 JUL-2000; 2000US-0614150.
 XX

(PEKE) PE CORP NY.
 Venter JC, Adams M, Li PWD, Myers EW;
 WPI: 2001-656860/75.
 P-PSDB; ABB64429.
 New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions
 Claim 1; SEQ ID NO 20078; 21pp + Sequence Listing; English.
 The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL01840-ABL16175) and the encoded proteins
 (ABB57737-ABB72072).
 The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 8844 BP; 2046 A; 2085 C; 2096 G; 2617 T; 0 other;

Query Match 2.4%; Score 18; DB 23; Length 8844;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 533 CCAACACGACATCTCACTT 550
 DB 7058 CCAACACGACATCTCACTT 7041
 |||||
 RESULT 56
 AAF28549/C
 ID AAF28549 standard; DNA; 92407 BP.
 XX
 AC AAF28549;
 XX
 DT 04-APR-2001 (first entry)
 XX
 DE Genomic fragment #36.
 XX
 KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200078968-A2.
 XX
 PD 28-DEC-2000.
 XX
 XX 16-JUN-2000; 2000WO-US16649.
 XX
 PR 18-JUN-1999; 99US-0140121.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lagace RE, Patterson C, Berg KL;
 XX
 DR WPI: 2001-041427/05.
 XX
 XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids
 XX
 PS Claim 1; Page 369-391; 545pp; English.
 XX

CC The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAR28514-AA28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.

XX Sequence 92407 BP: 26788 A: 17581 C: 20150 G: 27888 T: 0 other;
 SQ

Query Match 2.4%; Score 18; DB 22; Length 92407;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 ACCCAATCATCAACAG 586

|||||

DB 89597 AGCCAAATCATCAACAG 89580

RESULT 57

ABA51549/C

ID ABA51549 standard; DNA: 249 BP.

XX

AC ABA51549;

XX

DT 01-FEB-2002 (first entry)

XX

DE Human breast cell single exon nucleic acid probe #10244.

XX

KW Human; microarray; single exon probe; gene expression; breast;

XX

KW disease; cancer; ss.

XX

OS Homo sapiens.

XX

PN WO200157271-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00662.

XX

PR 04-FEB-2000; 2000US-0180312.

XX

PR 26-MAY-2000; 2000US-0207456.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687.

XX

PR 27-SEP-2000; 2000US-0236359.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI: 2001-496933/54.

XX

XX New spatially-addressable set of single exon nucleic acid probes,

XX

PT useful for measuring gene expression in sample derived from human

XX

PT breast, comprises number of single exon nucleic acid probes

XX

XX Claim 4; SEQ ID NO 10244: 327pp + sequence listing; English.

XX

XX The invention relates to a spatially-addressable set of single exon

XX

CC nucleic acid probes for measuring gene expression in a sample derived

XX

CC from human breast and Br 474 cells. The method involves contacting

XX

CC the probes with a collection of detectably labelled nucleic acids

XX

CC derived from mRNA of human breast, and then measuring the label

XX

CC bound to each probe of the microarray. The probes are useful for

XX

CC verifying the expression of regions of genomic DNA predicted to

CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 249 BP: 46 A; 76 C; 54 G; 73 T; 0 other;
 SQ

Query Match 2.3%; Score 17; DB 22; Length 249;

Best Local Similarity 100.0%; Pred. No. 28-02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 GATGTCAAAGCTGAACC 348

|||||

DB 170 GATGTCAAAGCTGAACC 154

RESULT 58

ABA36520/C

ID ABA36520 standard; DNA: 249 BP.

XX

AC ABA36520;

XX

DT 23-JAN-2002 (first entry)

XX

DE Probe #14986 for gene expression analysis in human heart cell sample.

XX

KW Human; gene expression; heart; microarray; vascular system; probe;

XX

KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX

KW congenital heart disease; ss.

XX

OS Homo sapiens.

XX

PN WO200157274-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00666.

XX

PR 04-FEB-2000; 2000US-0180312.

XX

PR 26-MAY-2000; 2000US-0207456.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687.

XX

PR 27-SEP-2000; 2000US-0236359.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI: 2001-488899/53.

XX

XX Single exon nucleic acid probes for analyzing gene expression in human

XX

PT hearts

XX

PS Claim 4; SEQ ID NO 14986: 530pp; English.

XX

XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 249 BP; 46 A; 76 C; 54 G; 73 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 249;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 GATGTCAAAAGCTGAACC 348

DB 170 GATGTCAAAAGCTGAACC 154

RESULT 59

AAC56190

ID AAC56190 standard; DNA: 350 BP.

XX AAC56190;

XX 25-JAN-2001 (first entry)

XX Eucalyptus grandis transcription factor DNA sequence #321.

XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.

XX Eucalyptus grandis.

OS WO200053724-A2.

XX 14-SEP-2000.

XX 09-MAR-2000; 2000WO-US06112.

XX 11-MAR-1999; 99US-0266513.

XX 18-AUG-1999; 99US-0149485.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Wood M, McGrath A, Shenk MA, Glenn M;

XX WPI; 2000-579369/54.

XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide

XX Claim 1; Page 129; 747pp; English.

XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

XX Sequence 350 BP; 66 A; 116 C; 111 G; 57 T; 0 other;

Query Match

2.3%; Score 17; DB 21; Length 350;

Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 0;

QY 661 CCGCATCTTCATCCACT 677

DB 244 ccgcattcttccact 260

RESULT 60

AAI80514/C

ID AAI80514 standard; cDNA: 395 BP.

XX AAI80514;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 574.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.

XX Homo sapiens.

OS WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac MT;

XX WPI; 2001-514838/56.

XX P-PSDB; AAO00583.

XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders.

XX Claim 1; SEQ ID NO 574; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activator/inhibitor activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 395 BP; 59 A; 164 C; 65 G; 106 T; 1 other;

Query Match

2.3%; Score 17; DB 22; Length 395;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 631 CCTGCTGCAGAGGAGA 647

DB 358 CCTGCTGCAGAGGAGA 342

RESULT 61
ABA46449/C
ID ABA46449 standard; DNA: 454 BP.
XX
AC ABA46449;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #5144.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
PA
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
DR

XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 1; SEQ ID NO 5144; 327pp + sequence listing: English.

XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and Br 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO.int/pub/published_pct_sequences.

XX
SQ Sequence 454 BP; 96 A; 135 C; 89 G; 134 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 2e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 GATGTCAAAGCTGAACC 348
|||||
DB 444 GATGTCAAAGCTGAACC 428

RESULT 62
ABA26652/C
ID ABA26652 standard; DNA: 454 BP.
XX
AC ABA26652;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #5118 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
PA
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
DR

XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX

PS Claim 1; SEQ ID No 5118; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 454 BP; 96 A; 135 C; 89 G; 134 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 2e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 GATGTCAAAGCTGAACC 348
|||||
DB 444 GATGTCAAAGCTGAACC 428

RESULT 63
AAH77208/C
ID AAH77208 standard; cDNA: 461 BP.
XX
AC AAH77208;
XX
DT 30-JAN-2002 (first entry)

```

XX Arthropod defensin cDNA from Scolopendra canidens Ds clone ascl.pk03.
DE
XX
XX Arthropod; defensin; ascl.pk03; antimicrobial; microbial pathogen;
KW gene therapy; ss.
XX
XX Scolopendra canidens Ds.
OS
XX Key Location/Qualifiers
FH 50..235
FT CDS /*tag=
FT /product= "Defensin"
XX
XX EP1146052-A2.
PN
XX
XX 17-OCT-2001.
PD
XX
XX 12-APR-2001; 2001EP-0303488.
PF
XX
XX 14-APR-2000; 2000US-197279p.
PR
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA
XX Presnall JK, Weng Z, Wong JF;
PI
XX WPI: 2002-001056/01.
DR
XX P-PSDB: AAG77931.
XX
XX New polynucleotides encoding Arthropod defensin polypeptides for
PT improving the microbial resistance of plants and animals by gene
PT therapy.
XX
XX Claim 6; Page 17; 22pp; English.
PS
XX The sequence represents a novel Arthropod defensin from S. canidens Ds
CC clone ascl.pk03. The invention relates to novel polynucleotides encoding
CC Arthropod defensin polypeptides. The polypeptides and polynucleotides of
CC the invention have antimicrobial activity. The polynucleotides may be
CC used to genetically engineer cells and organisms, especially plants, to
CC alter their resistance to a wide range of microbial pathogens. The
CC polynucleotides may also have a use in gene therapy.
XX
XX Sequence 461 BP; 143 A; 73 C; 85 G; 157 T; 3 other;
SQ
Query Match 2.3%; Score 17; DB 24; Length 461;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 194 CTTCAAAATCCAGTCC 210
DB 211 CTTCAAAATCCAGTGC 195
RESULT 64
ABA63465/C
ID ABA63465 standard; DNA: 570 BP.
XX
XX ABA63465;
AC
XX
XX 01-FEB-2002 (first entry)
DT
XX
XX Human foetal liver single exon nucleic acid probe #11770.
DE
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KW
XX Homo sapiens.
OS
XX
XX WO200157277-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00669.
PF

```

```

XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-483447/52.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
PT
XX Claim 1; SEQ ID NO 11770; 639pp + sequence listing; English.
PS
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;
SQ
Query Match 2.3%; Score 17; DB 22; Length 570;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 550 TTCCTGCTGGACAAAC 566
DB 167 TTCCTGCTGGACAAAC 151
RESULT 65
ABA30663/C
ID ABA30663 standard; DNA: 570 BP.
XX
XX ABA30663;
AC
XX
XX 23-JAN-2002 (first entry)
DT
XX
XX Probe #9129 for gene expression analysis in human heart cell sample.
DE
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157274-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA

```

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI: 2001-488899/53.
 XX PT Single exon nucleic acid probes for analyzing gene expression in human
 XX PT hearts -
 XX PS Claim 1: SEQ ID No. 9129; 530pp; English.
 XX CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC CC Note: the sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX CC Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;
 XX SQ

Query Match 2.3%; Score 17; DB 22; Length 570;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 TTCCTTGCTGGACAAAC 566
 Db 167 TTCCTTGCTGGACAAAC 151
 |||||

RESULT 66
 AAK11998/C
 ID AAK11998 standard; DNA; 570 BP.
 XX AC AAK11998;
 XX DT 05-NOV-2001 (first entry)
 XX DE Human brain expressed single exon probe SEQ ID NO: 11989.
 XX KW Human: brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 XX OS Homo sapiens.
 XX PN WO200157275-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00667.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI: 2001-483446/52.
 XX PT Single exon nucleic acid probes for analyzing gene expression in human
 XX PT brains -

XX PS Example 4: SEQ ID NO: 11989; 650pp + Sequence Listing; English.
 XX CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX CC Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;
 XX SQ

Query Match 2.3%; Score 17; DB 22; Length 570;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 TTCCTTGCTGGACAAAC 566
 Db 167 TTCCTTGCTGGACAAAC 151
 |||||

RESULT 67
 AAK37702/C
 ID AAK37702 standard; DNA; 570 BP.
 XX AC AAK37702;
 XX DT 06-NOV-2001 (first entry)
 XX DE Human bone marrow expressed single exon probe SEQ ID NO: 12259.
 XX KW Human: bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX OS Homo sapiens.
 XX PN WO200157276-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00668.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI: 2001-488900/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human bone marrow -
 XX PS Example 4: SEQ ID NO: 12259; 658pp + Sequence Listing; English.
 XX CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX CC Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;
 XX SQ

```

Query Match      2.3%; Score 17; DB 22; Length 570;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 TTCCTTGCTGCACAAAC 566
DB 167 TTCCTTGCTGCACAAAC 151

RESULT 68
AAH18460/C
ID AAH18460 standard; DNA; 570 BP.
XX AC AAH18460;
XX DT 12-OCT-2001 (first entry)
XX DE Probe #8393 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US006070.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX PS Claim 25; SEQ ID NO 8393; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX CC from human HeLa cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;

Query Match      2.3%; Score 17; DB 22; Length 570;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 TTCCTTGCTGCACAAAC 566
DB 167 TTCCTTGCTGCACAAAC 151

RESULT 69
AAH18460/C
ID AAH18460 standard; DNA; 570 BP.
XX AC AAH18460;
XX DT 12-OCT-2001 (first entry)
XX DE Probe #8393 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US006070.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX PS Claim 25; SEQ ID NO 8393; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX CC from human HeLa cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;

Query Match      2.3%; Score 17; DB 22; Length 570;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 TTCCTTGCTGCACAAAC 566
DB 167 TTCCTTGCTGCACAAAC 151

RESULT 69
AAH18460/C
ID AAH18460 standard; DNA; 570 BP.
XX AC AAH18460;
XX DT 12-OCT-2001 (first entry)
XX DE Probe #8393 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US006070.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-488901/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID NO 12263; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;

Query Match      2.3%; Score 17; DB 22; Length 570;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 550 TTCCTTGCTGCACAAAC 566
DB 167 TTCCTTGCTGCACAAAC 151

RESULT 70
AAH04794
ID AAH04794 standard; cDNA; 751 BP.
XX AC AAH04794;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:1629.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX KW Homo sapiens.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX SQ Sequence 570 BP; 144 A; 142 C; 136 G; 148 T; 0 other;

```

PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 94JP-0248036.
 PR 27-AUG-1999; 98JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 PS Claim 1: SEQ ID 1629; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 751 BP; 160 A; 195 C; 199 G; 193 T; 4 other;

Query Match 2.3%; Score 17; DB 22; Length 751;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 619 CACTGTATCTCTCCCTGC 635
 Db 236 cactgtatctctccctgc 252
 |||||

RESULT 71
 AAL25025/c
 ID AAL25025 standard; cDNA: 775 BP.
 XX
 AC AAL25025;
 XX
 DT 07-DEC-2001 (first entry)
 XX
 DE Human breast cancer expressed polynucleotide 17482.
 XX
 DE Human; breast cancer; cell marker; cytostatic; ss.
 KW Human; breast cancer; cell marker; cytostatic; ss.
 XX
 OS Homo sapiens.
 XX
 PI Dillon DC, Xu J;

PN WO200151628-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 10-JAN-2001; 2001WO-US00798.
 XX
 PR 14-JAN-2000; 2000US-0176077.
 PR 14-MAR-2000; 2000US-0189167.
 PR 24-MAR-2000; 2000US-0192099.
 PR 29-MAR-2000; 2000US-0193480.
 PR 15-MAY-2000; 2000US-0205230.
 PR 09-JUN-2000; 2000US-0211315.
 PR 25-JUL-2000; 2000US-0220534.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Lillie J, Xu Y, Wang Y, Steinmann K;
 XX
 DR WPI; 2001-451856/48.
 XX
 PT New peptide useful as a marker for the diagnosis of breast cancer
 XX
 PS Claim 1: Page 3228; 3695pp; English.
 XX
 CC The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 XX
 SQ Sequence 775 BP; 188 A; 179 C; 186 G; 222 T; 0 other;

Query Match 2.3%; Score 17; DB 22; Length 775;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 AGGAAGCTCAGAAATTG 253
 Db 146 AGGAAGCTCAGAAATTG 130
 |||||

RESULT 72
 AAV61275/c
 ID AAV61275 standard; cDNA: 789 BP.
 XX
 AC AAV61275;
 XX
 DT 06-JAN-1999 (first entry)
 XX
 DE 3' cDNA sequence of prostate tumour clone J1-21.
 XX
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9837093-A2.
 XX
 PD 27-AUG-1998.
 XX
 PF 25-FEB-1998; 98WO-US03492.
 XX
 PR 09-FEB-1998; 98US-0020956.
 PR 25-FEB-1997; 97US-0806099.
 PR 01-AUG-1997; 97US-0904804.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Dillon DC, Xu J;

XX WPI: 1998-609886/51.
 XX polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 XX
 XX Claim 12; Page 48; 130pp; English.
 XX The present sequence is a DNA which encodes an immunogenic portion
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 XX
 XX Sequence 789 BP; 129 A; 240 C; 159 G; 195 T; 66 other;
 SQ

Query Match 2.3%; Score 17; DB 19; Length 789;
 Best Local Similarity 100.0%; Pred. No. 2e-02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 AGCCTGTTCCCTCCTG 514
 DB 131 AGCCTGTTCCCTCCTG 115
 |||||

RESULT 73
 AAV58509/c
 ID AAV58509 standard; cDNA: 789 BP.
 XX
 AC AAV58509;
 XX
 XX 08-DEC-1998 (first entry)
 DT
 DE 3' fragment of prostate tumour specific gene J1-21.
 XX
 XX Prostate tumour specific gene; human; prostate cancer; detection;
 KW therapy; ss.
 KW Homo sapiens.
 OS
 XX
 XX WO9837418-A2.
 PN
 XX
 XX 27-AUG-1998.
 PD
 XX
 XX 25-FEB-1998; 98WO-US03690.
 PF
 XX
 XX 09-FEB-1998; 98US-0904809.
 PR
 XX 23-FEB-1997; 97US-0806596.
 PR
 XX 01-AUG-1997; 97US-0904809.
 PR
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX Dillon DC, Xu J;
 PI
 XX
 XX WPI: 1998-480805/41.
 DR
 XX
 XX Novel human prostate specific tumour protein and fragments - useful
 PT for detecting and treating prostate cancers
 PT
 XX
 XX Claim 1; Page 50; 141pp; English.
 PS
 XX
 XX This sequence represents a human prostate tumour specific gene, and can
 CC be used in the method of the invention. The method is for detecting
 CC prostate cancer comprising contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 CC
 XX

SQ Sequence 789 BP; 129 A; 240 C; 159 G; 195 T; 66 other;
 Query Match 2.3%; Score 17; DB 19; Length 789;
 Best Local Similarity 100.0%; Pred. No. 2e-02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 AGCCTGTTCCCTCCTG 514
 DB 131 AGCCTGTTCCCTCCTG 115
 |||||

RESULT 74
 AAA06272/c
 ID AAA06272 standard; cDNA: 789 BP.
 XX
 AC AAA06272;
 XX
 XX 13-JUN-2000 (first entry)
 DT
 DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:32.
 XX
 XX Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
 KW immunogenic; cytostatic; vaccine; ss.
 KW
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200004149-A2.
 PN
 XX
 XX 27-JAN-2000.
 PD
 XX
 XX 14-JUL-1999; 99WO-US15838.
 PF
 XX
 XX 14-JUL-1998; 98US-0115453.
 PR
 XX 14-JUL-1998; 98US-0116134.
 PR
 XX 23-SEP-1998; 98US-0159812.
 PR
 XX 23-SEP-1998; 98US-0159822.
 PR
 XX 15-JAN-1999; 99US-0232149.
 PR
 XX 15-JAN-1999; 99US-0232880.
 PR
 XX 09-APR-1999; 99US-0288946.
 PR
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
 PI
 XX
 XX WPI: 2000-171268/15.
 DR
 XX
 XX New polypeptide useful for treating and diagnosing prostate cancer
 PT comprises an immunogenic portion of prostate tumor protein -
 PT
 XX
 XX Claim 50; Page 110; 263pp; English.
 PS
 XX
 XX The present invention describes isolated polypeptides, comprising an
 CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
 CC and polynucleotides encoding them have cytostatic activity and can be
 CC used in vaccines and in gene therapy. The polypeptides and
 CC polynucleotides encoding them, antigen presenting cells which express
 CC the polypeptides, antibodies against the polypeptides and vaccines
 CC comprising them can be used for inhibiting the development of prostate
 CC cancer in a patient. The polypeptides can be used to generate antibodies
 CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
 CC the polynucleotides encoding the polypeptides can be used as a probe or
 CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
 CC AAY82000 to AAY82020 represent sequences used in the exemplification of
 CC the present invention.
 CC
 XX
 XX Sequence 789 BP; 129 A; 240 C; 159 G; 195 T; 66 other;
 SQ

Query Match 2.3%; Score 17; DB 21; Length 789;
 Best Local Similarity 100.0%; Pred. No. 2e-02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Wed Aug 7 05:46:15 2002

Search completed: August 6, 2002, 19:25:39
Job time: 7033 sec

```
QY 498 AGCCTGTTGCCGTCCTG 514
DB 131 AGCCTGTTGCCGTCCTG 115
|||||
|||||

RESULT 75
AAS63480/c
ID AAS63480 standard: cDNA: 789 BP.
AC AAS63480;
XX
XX
DT 29-JAN-2002 (first entry)
XX
DE Human prostate cDNA sequence #32.
XX
XX Human prostate cancer; ss: cytostatic; immunostimulant; tumour.
XX
XX Homo sapiens.
XX
XX WO200173032-A2.
XX
XX 04-OCT-2001.
XX
XX 27-MAR-2001; 2001WO-US09919.
XX
XX 27-MAR-2000; 2000US-0536857.
XX
XX 09-MAY-2000; 2000US-0568100.
XX
XX 12-MAY-2000; 2000US-0570737.
XX
XX 13-JUN-2000; 2000US-0593793.
XX
XX 27-JUN-2000; 2000US-0605783.
XX
XX 10-AUG-2000; 2000US-0636215.
XX
XX 29-AUG-2000; 2000US-0651236.
XX
XX 06-SEP-2000; 2000US-0657279.
XX
XX 02-OCT-2000; 2000US-0679426.
XX
XX 10-OCT-2000; 2000US-0685166.
XX
XX (CORI-) CORIXA COMP.
XX
XX Xu J, Dillon DC, Harlocker SL, Jiang Y, Kalos MD;
XX Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
XX Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX WPI; 2001-639232/73.
XX
XX New human prostate-specific polypeptides and polynucleotides useful for
XX the diagnosis and treatment of cancer, especially prostate cancer -
XX
XX Claim 1; Page 242; 579pp; English.
XX
XX The invention relates to isolated prostate-specific
XX polynucleotides, polypeptides, fusion proteins of the polypeptides,
XX antibodies raised against the polypeptides (or antigenic epitopes
XX derived from them) and antigen-presenting cells expressing the
XX polypeptides. The antibodies are useful for detecting the presence of
XX cancer, especially prostate cancer. The polypeptides, polynucleotides and
XX the antigen-presenting cells are useful for stimulating and/or expanding
XX T cells specific for a tumour protein, and for inhibiting the development
XX of cancer, especially prostate cancer. Compositions comprising the
XX polynucleotide and/or polypeptide are useful for stimulating an immune
XX response, and for treating cancer. The oligonucleotide is useful for
XX detecting cancer. The present sequence is a prostate specific
XX polynucleotide of the invention.
XX
XX Sequence 789 BP; 129 A; 240 C; 159 G; 195 T; 66 other;

Query Match 2.3%; Score 17; DB 22; Length 789;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 AGCCTGTTGCCGTCCTG 514
DB 131 AGCCTGTTGCCGTCCTG 115
|||||
|||||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 18:16:47 ; Search time 49.1 Seconds

(without alignments)
3732.028 Million cell updates/sec

Title: US-10-020-139-1_COPY_48_793

Perfect score: 746
Sequence: 1 GATCCTTCAGCTTTGGAAC.....AACCCAGCTGCAACCTCA 746

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size: 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 250 summaries

Database: Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*\n2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*\n3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*\n4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*\n5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*\n6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	2.7	4692	US-08-916-917-1	Sequence 1, Appli
2	20	2.7	4692	US-08-972-631-1	Sequence 1, Appli
3	20	2.7	4692	US-08-972-629-1	Sequence 1, Appli
4	20	2.7	4692	US-08-972-630-1	Sequence 1, Appli
5	20	2.7	4692	US-08-972-211-1	Sequence 1, Appli
6	20	2.7	4692	US-08-925-170-1	Sequence 1, Appli
7	18	2.4	2345	US-08-955-937A-1	Sequence 1, Appli
8	18	2.4	2345	US-09-300-985-1	Sequence 1, Appli
9	17	2.3	789	US-09-020-956-32	Sequence 32, Appl
10	17	2.3	789	US-09-030-607-32	Sequence 32, Appl
11	17	2.3	789	US-09-439-313-32	Sequence 32, Appl
12	17	2.3	4403765	US-09-103-840A-2	Sequence 2, Appli
13	17	2.3	4411529	US-09-103-840A-1	Sequence 4, Appli
14	16	2.1	396	US-09-534-638-4	Sequence 142, App
15	16	2.1	601	US-09-328-111-142	Sequence 87, Appl
16	16	2.1	735	US-09-327-357-87	Sequence 42, Appl
17	16	2.1	1152	US-08-844-188-42	Sequence 4, Appli
18	16	2.1	1493	US-08-233-389C-4	Sequence 4, Appli
19	16	2.1	1493	US-08-801-863-4	Sequence 4, Appli
20	16	2.1	1493	US-08-486-596A-4	Sequence 4, Appli
21	16	2.1	1493	US-09-004-713-4	Sequence 4, Appli
22	16	2.1	2057	US-09-059-522-2	Sequence 2, Appli
23	16	2.1	2057	US-09-382-027-2	Sequence 2, Appli
24	16	2.1	2132	US-08-844-188-39	Sequence 39, Appl
25	16	2.1	2478	US-08-481-626-1	Sequence 1, Appli
26	16	2.1	2589	US-08-653-740-6	Sequence 6, Appli
27	16	2.1	2589	US-09-073-594-6	Sequence 6, Appli

28	16	2.1	2589	US-09-275-925-6	Sequence 6, Appli
29	16	2.1	3155	US-09-442-100-7	Sequence 7, Appli
30	16	2.1	3256	US-08-968-751-3	Sequence 3, Appli
31	16	2.1	3565	US-08-306-691B-44	Sequence 44, Appl
32	16	2.1	3565	PCT-US93-06251-20	Sequence 20, Appl
33	16	2.1	4020	US-09-050-159-130	Sequence 130, App
34	16	2.1	4024	US-09-162-484-18	Sequence 18, Appl
35	16	2.1	8535	US-08-716-351A-1	Sequence 1, Appli
36	16	2.1	9840	US-09-534-638-1	Sequence 183, App
37	16	2.1	37950	US-09-338-907-183	Sequence 183, App
38	16	2.1	37950	US-09-218-207-183	Sequence 2, Appli
39	16	2.1	4403765	US-09-103-840A-2	Sequence 2, Appli
40	15	2.0	46	US-08-373-124A-128	Sequence 128, App
41	15	2.0	46	US-08-435-628-128	Sequence 128, App
42	15	2.0	50	US-08-998-099-289	Sequence 289, App
43	15	2.0	81	US-08-477-928A-43	Sequence 43, Appl
44	15	2.0	227	US-08-943-731-12	Sequence 12, Appl
45	15	2.0	378	US-09-240-274-189	Sequence 189, App
46	15	2.0	456	US-08-557-309B-16	Sequence 16, Appl
47	15	2.0	456	US-08-834-306-16	Sequence 16, Appl
48	15	2.0	456	US-08-993-674A-16	Sequence 16, Appl
49	15	2.0	495	US-09-323-472A-5	Sequence 5, Appli
50	15	2.0	520	US-07-743-518-25	Sequence 25, Appl
51	15	2.0	553	US-08-721-488-7	Sequence 7, Appli
52	15	2.0	571	US-08-803-899-1	Sequence 1, Appli
53	15	2.0	601	US-09-385-982-231	Sequence 231, App
54	15	2.0	728	US-08-398-416-604	Sequence 604, App
55	15	2.0	877	US-08-173-510B-95	Sequence 95, Appl
56	15	2.0	877	US-08-458-218-93	Sequence 93, Appl
57	15	2.0	877	US-08-450-497-95	Sequence 95, Appl
58	15	2.0	984	US-09-446-504-4	Sequence 4, Appli
59	15	2.0	984	US-09-712-266-4	Sequence 4, Appli
60	15	2.0	1192	US-09-142-565-1	Sequence 1, Appli
61	15	2.0	1316	US-08-579-667-3	Sequence 3, Appli
62	15	2.0	1320	US-08-841-038A-1	Sequence 1, Appli
63	15	2.0	1320	US-09-059-178-1	Sequence 1, Appli
64	15	2.0	1369	US-08-642-541-1	Sequence 1, Appli
65	15	2.0	1369	US-08-642-541-3	Sequence 3, Appli
66	15	2.0	1369	US-09-260-889-1	Sequence 1, Appli
67	15	2.0	1369	US-09-260-889-3	Sequence 3, Appli
68	15	2.0	1377	US-08-174-467-20	Sequence 20, Appl
69	15	2.0	1377	US-08-452-071-20	Sequence 20, Appl
70	15	2.0	1415	US-08-961-083-157	Sequence 157, App
71	15	2.0	1423	US-07-829-954-1	Sequence 1, Appli
72	15	2.0	1423	US-07-994-423-1	Sequence 1, Appli
73	15	2.0	1423	US-08-421-891-1	Sequence 1, Appli
74	15	2.0	1434	US-09-040-799-2	Sequence 2, Appli
75	15	2.0	1457	US-09-444-053-3	Sequence 3, Appli
76	15	2.0	1528	US-08-477-928A-1	Sequence 1, Appli
77	15	2.0	1533	PCT-US91-04274A-1	Sequence 1, Appli
78	15	2.0	1543	US-09-227-357-26	Sequence 26, Appl
79	15	2.0	1654	US-09-416-050A-3	Sequence 3, Appli
80	15	2.0	1654	US-09-664-800-3	Sequence 3, Appli
81	15	2.0	1654	US-09-665-309-3	Sequence 3, Appli
82	15	2.0	1654	US-08-661-569-3	Sequence 3, Appli
83	15	2.0	1676	US-08-936-165A-212	Sequence 212, App
84	15	2.0	1689	PCT-US91-04274A-2	Sequence 2, Appli
85	15	2.0	1742	US-08-477-928A-44	Sequence 44, Appl
86	15	2.0	1785	US-08-307-485A-6	Sequence 6, Appli
87	15	2.0	1785	US-08-465-809-1	Sequence 1, Appli
88	15	2.0	1794	US-08-427-497E-5	Sequence 5, Appli
89	15	2.0	1814	PCT-US96-00547-54	Sequence 54, Appl
90	15	2.0	1836	US-07-929-580B-1	Sequence 1, Appli
91	15	2.0	1839	US-07-929-580B-4	Sequence 4, Appli
92	15	2.0	1845	US-08-652-207A-1	Sequence 1, Appli
93	15	2.0	1845	US-09-488-857B-3	Sequence 3, Appli
94	15	2.0	1892	PCT-US96-00547-41	Sequence 41, Appl
95	15	2.0	1908	US-09-428-584-3	Sequence 3, Appli
96	15	2.0	1927	US-09-040-799-1	Sequence 1, Appli
97	15	2.0	1990	US-09-255-911-1	Sequence 1, Appli
98	15	2.0	2037	US-08-840-767-9	Sequence 9, Appli
99	15	2.0	2310	US-08-261-822A-4	Sequence 4, Appli
100	15	2.0	2310	PCT-US95-07744A-4	Sequence 4, Appli

101	15	2.0	2466	1	US-07-921-796-5	Sequence 5, Appl	174	15	2.0	4517	4	US-09-140-804-9	Sequence 9, Appl
102	15	2.0	2509	2	US-09-014-969-1	Sequence 1, Appl	c 175	15	2.0	4870	4	US-08-664-962B-7	Sequence 7, Appl
103	15	2.0	2540	4	US-08-838-151A-18	Sequence 18, Appl	c 176	15	2.0	4870	4	US-09-311-743-7	Sequence 7, Appl
104	15	2.0	2609	4	US-09-141-212-7	Sequence 7, Appl	c 177	15	2.0	5113	3	US-08-973-273-2	Sequence 2, Appl
105	15	2.0	2609	4	US-09-561-138-7	Sequence 7, Appl	c 178	15	2.0	5191	1	US-08-340-428B-1	Sequence 1, Appl
106	15	2.0	2625	4	US-09-378-255-1	Sequence 1, Appl	c 179	15	2.0	5191	5	PCT-US93-07306-1	Sequence 1, Appl
107	15	2.0	2625	4	US-09-715-336-1	Sequence 1, Appl	c 180	15	2.0	5408	1	US-08-441-139-15	Sequence 15, Appl
108	15	2.0	2634	2	US-08-818-514-1	Sequence 1, Appl	c 181	15	2.0	5449	4	US-09-546-990-1	Sequence 1, Appl
109	15	2.0	2634	2	US-08-818-514-2	Sequence 2, Appl	c 182	15	2.0	5559	1	US-08-287-442-3	Sequence 3, Appl
110	15	2.0	2634	3	US-09-115-934A-1	Sequence 1, Appl	c 183	15	2.0	5559	1	US-08-459-701-3	Sequence 3, Appl
111	15	2.0	2634	3	US-09-115-934A-2	Sequence 2, Appl	c 184	15	2.0	5559	1	US-08-480-298-3	Sequence 3, Appl
112	15	2.0	2669	4	US-09-141-212-9	Sequence 9, Appl	c 185	15	2.0	5559	1	US-08-459-701-3	Sequence 3, Appl
113	15	2.0	2669	4	US-09-561-138-9	Sequence 9, Appl	c 186	15	2.0	5559	1	US-08-480-298-3	Sequence 3, Appl
114	15	2.0	2693	4	US-09-141-212-5	Sequence 5, Appl	c 187	15	2.0	5648	5	PCT-US96-03940-1	Sequence 1, Appl
115	15	2.0	2693	4	US-09-561-138-5	Sequence 5, Appl	c 188	15	2.0	9045	3	US-09-121-321-1	Sequence 1, Appl
116	15	2.0	2709	4	US-09-251-372-1	Sequence 1, Appl	c 189	15	2.0	9045	4	US-08-933-803A-1	Sequence 1, Appl
117	15	2.0	2709	4	US-09-811-241-1	Sequence 1, Appl	c 190	15	2.0	9763	3	US-08-973-273-1	Sequence 1, Appl
118	15	2.0	2820	5	PCT-US93-11723-1	Sequence 1, Appl	c 191	15	2.0	10763	2	US-08-973-273-1	Sequence 1, Appl
119	15	2.0	2855	2	US-08-776-597A-1	Sequence 1, Appl	c 192	15	2.0	10763	2	US-08-973-273-1	Sequence 1, Appl
120	15	2.0	2855	2	US-08-693-228-1	Sequence 1, Appl	c 193	15	2.0	10952	2	US-08-602-036A-1	Sequence 1, Appl
121	15	2.0	2940	3	US-09-195-868-11	Sequence 11, Appl	c 194	15	2.0	10952	2	US-08-502-374A-1	Sequence 1, Appl
122	15	2.0	2944	2	US-08-696-944-18	Sequence 18, Appl	c 195	15	2.0	10952	2	US-08-502-374A-1	Sequence 1, Appl
123	15	2.0	3000	2	US-08-928-692-9	Sequence 9, Appl	c 196	15	2.0	12311	4	US-08-750-717-1	Sequence 1, Appl
124	15	2.0	3006	4	US-09-552-351-1	Sequence 1, Appl	c 197	15	2.0	13011	2	US-08-791-849A-14	Sequence 14, Appl
125	15	2.0	3006	4	US-09-552-351-1	Sequence 1, Appl	c 198	15	2.0	13011	2	US-08-943-731-1	Sequence 1, Appl
126	15	2.0	3040	1	US-08-446-794A-1	Sequence 1, Appl	c 199	15	2.0	18609	4	US-08-814-095-7	Sequence 7, Appl
127	15	2.0	3040	2	US-08-750-007-2	Sequence 2, Appl	c 200	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
128	15	2.0	3040	2	US-08-945-024-1	Sequence 1, Appl	c 201	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
129	15	2.0	3040	2	US-09-378-255-3	Sequence 3, Appl	c 202	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
130	15	2.0	3040	2	US-09-378-255-3	Sequence 3, Appl	c 203	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
131	15	2.0	3089	1	US-07-921-796-7	Sequence 7, Appl	c 204	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
132	15	2.0	3097	2	US-08-599-455B-1	Sequence 1, Appl	c 205	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
133	15	2.0	3097	2	US-09-069-781B-1	Sequence 1, Appl	c 206	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
134	15	2.0	3132	2	US-08-224-482-3	Sequence 3, Appl	c 207	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
135	15	2.0	3132	3	US-09-205-921-1	Sequence 1, Appl	c 208	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
136	15	2.0	3255	3	US-08-916-917-11	Sequence 11, Appl	c 209	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
137	15	2.0	3255	3	US-09-325-170-11	Sequence 11, Appl	c 210	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
138	15	2.0	3375	1	US-08-381-931B-1	Sequence 1, Appl	c 211	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
139	15	2.0	3387	1	US-08-599-455B-1	Sequence 1, Appl	c 212	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
140	15	2.0	3387	1	US-08-599-455B-1	Sequence 1, Appl	c 213	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
141	15	2.0	3387	1	US-08-599-455B-1	Sequence 1, Appl	c 214	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
142	15	2.0	3574	4	US-09-446-504-83	Sequence 83, Appl	c 215	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
143	15	2.0	3574	4	US-09-712-266-83	Sequence 83, Appl	c 216	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
144	15	2.0	3620	4	US-09-446-504-55	Sequence 55, Appl	c 217	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
145	15	2.0	3620	4	US-09-712-266-55	Sequence 55, Appl	c 218	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
146	15	2.0	3630	4	US-09-378-255-5	Sequence 5, Appl	c 219	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
147	15	2.0	3630	4	US-09-141-212-1	Sequence 1, Appl	c 220	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
148	15	2.0	3630	4	US-09-141-212-3	Sequence 3, Appl	c 221	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
149	15	2.0	3630	4	US-09-552-351-3	Sequence 3, Appl	c 222	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
150	15	2.0	3630	4	US-09-251-372-5	Sequence 5, Appl	c 223	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
151	15	2.0	3630	4	US-09-561-138-1	Sequence 1, Appl	c 224	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
152	15	2.0	3630	4	US-09-561-138-3	Sequence 3, Appl	c 225	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
153	15	2.0	3630	4	US-09-715-336-5	Sequence 5, Appl	c 226	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
154	15	2.0	3630	4	US-09-811-241-5	Sequence 5, Appl	c 227	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
155	15	2.0	3630	4	US-09-802-839-3	Sequence 3, Appl	c 228	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
156	15	2.0	3632	1	US-08-424-788-4	Sequence 4, Appl	c 229	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
157	15	2.0	3632	1	US-08-110-683-1	Sequence 1, Appl	c 230	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
158	15	2.0	3632	2	US-08-683-743-1	Sequence 1, Appl	c 231	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
159	15	2.0	3632	2	US-08-477-166-1	Sequence 1, Appl	c 232	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
160	15	2.0	3632	2	US-08-477-166-1	Sequence 1, Appl	c 233	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
161	15	2.0	3632	2	US-08-477-166-1	Sequence 1, Appl	c 234	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
162	15	2.0	3651	4	PCT-US93-11638-1	Sequence 13, Appl	c 235	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
163	15	2.0	3666	6	5248670-3	Sequence 13, Appl	c 236	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
164	15	2.0	3691	3	US-09-195-868-12	Sequence 12, Appl	c 237	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
165	15	2.0	3705	5	PCT-US96-03940-7	Sequence 7, Appl	c 238	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
166	15	2.0	3854	2	US-08-599-455B-42	Sequence 42, Appl	c 239	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
167	15	2.0	3854	2	US-09-069-781B-42	Sequence 42, Appl	c 240	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
168	15	2.0	3876	5	PCT-US95-17026-1	Sequence 1, Appl	c 241	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
169	15	2.0	4092	2	US-08-469-537A-106	Sequence 106, App	c 242	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
170	15	2.0	4147	3	US-08-560-005-1	Sequence 1, Appl	c 243	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
171	15	2.0	4147	3	US-09-418-540-1	Sequence 1, Appl	c 244	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
172	15	2.0	4157	2	US-08-162-146-2	Sequence 2, Appl	c 245	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App
173	15	2.0	4157	2	US-09-314-127-2	Sequence 2, Appl	c 246	15	2.0	35060	3	US-09-338-907-179	Sequence 179, App

c 247 14 1.9 420 4 US-09-479-776-11 Sequence 11, Appl
 248 14 1.9 430 4 US-09-060-756-178 Sequence 178, App
 249 14 1.9 443 4 US-09-586-125-1 Sequence 1, Appl
 250 14 1.9 480 1 US-08-282-581-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
 US-08-916-917-1
 ; Sequence 1, Application US/08916917
 ; Patent No. 5856132
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Len
 ; APPLICANT: Hawkins, Phillip Thomas
 ; APPLICANT: Braselmann, Sylvia
 ; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 ; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3', KINASE
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds, LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10036-2811
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: FASTSEQ for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/916,917
 ; FILING DATE: 15-AUG-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/672,211
 ; FILING DATE: 27-JUN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Abrams, Samuel B.
 ; REGISTRATION NUMBER: 30,605
 ; REFERENCE/DOCKET NUMBER: 8549-0006-999
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-493-4935
 ; TELEFAX: 650-493-5556
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4692 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-916-917-1

Query Match 2.7%; Score 20; DB 2; Length 4692;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 665 ATCTTCATCCACTCCCTGGA 684
 |||||
 Db 2448 ATCTTCATCCACTCCCTGGA 2467

RESULT 2
 US-08-972-631-1
 ; Sequence 1, Application US/08972631
 ; Patent No. 5856133
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Len
 ; APPLICANT: Hawkins, Phillip T.
 ; TITLE OF INVENTION: G-BETA-GAMMA REGULATED

; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 2730 Sand Hill Road
 ; CITY: Menlo Park
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/972,631
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/672,211
 ; FILING DATE: 27-JUN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Halluin, Albert P.
 ; REGISTRATION NUMBER: 25,277
 ; REFERENCE/DOCKET NUMBER: 8549-0005-999
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415)854-3660
 ; TELEFAX: (415)854-3694
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4692 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: cdna
 ; US-08-972-631-1

Query Match 2.7%; Score 20; DB 2; Length 4692;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 665 ATCTTCATCCACTCCCTGGA 684
 |||||
 Db 2448 ATCTTCATCCACTCCCTGGA 2467

RESULT 3
 US-08-972-629-1
 ; Sequence 1, Application US/08972629
 ; Patent No. 5859201
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephens, Len
 ; APPLICANT: Hawkins, Phillip T.
 ; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
 ; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 2730 Sand Hill Road
 ; CITY: Menlo Park
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/972,629
 ; FILING DATE:
 ; CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-972-629-1

Query Match 2.7%; Score 20; DB 2; Length 4692;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTGGA 684
|||||
DB 2448 ATCTTCATCCACTCCCTGGA 2467

RESULT 4
US-08-972-630-1
Sequence 1, Application US/08972630
Patent No. 5869271
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip T.
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,630
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-972-630-1

Query Match 2.7%; Score 20; DB 2; Length 4692;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTGGA 684
|||||
DB 2448 ATCTTCATCCACTCCCTGGA 2467

RESULT 5
US-08-672-211-1
Sequence 1, Application US/08672211
Patent No. 5874273
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip T.
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/672,211
FILING DATE: 27-JUN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-672-211-1

Query Match 2.7%; Score 20; DB 2; Length 4692;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCCTGGA 684
|||||
DB 2448 ATCTTCATCCACTCCCTGGA 2467

RESULT 6
US-09-225-170-1
Sequence 1, Application US/09225170
Patent No. 6017763
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip Thomas
APPLICANT: Braselmann, Sylvia
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE

NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225.170
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916.917
FILING DATE: 15-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8549-0006-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5536
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4692 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-225-170-1

Query Match 2.7%: Score 20; DB 3; Length 4692;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 ATCTTCATCCACTCCTCTGCA 684
|||||
DB 2448 ATCTTCATCCACTCCTCTGCA 2467

RESULT 7
US-08-955-937A-1/c
Sequence 1, Application US/08955937A
Patent No. 6020161
GENERAL INFORMATION:
APPLICANT: WU, SHUJIAN
APPLICANT: SWEET, RAYMOND
APPLICANT: TRUNEH, ALEMSGED
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN
TITLE OF INVENTION: GENE SUPERFAMILY
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/955.937A
FILING DATE: 17-OCT-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/056,152
FILING DATE: 19-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70228
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2345 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-955-937A-1

Query Match 2.4%: Score 18; DB 3; Length 2345;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 GGCCCATCATTTGCCGAGA 428
|||||
DB 1986 GGCCCATCATTTGCCGAGA 1969

RESULT 8
US-09-300-985-1/c
Sequence 1, Application US/09300985A
Patent No. 6232441
GENERAL INFORMATION:
APPLICANT: WU, SHUJIAN
APPLICANT: SWEET, RAYMOND W.
APPLICANT: TRUNEH, ALEMSGED
APPLICANT: HORLE, MARK ROBERT
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN GENE
TITLE OF INVENTION: SUPERFAMILY
FILE REFERENCE: GH-70228-1
CURRENT APPLICATION NUMBER: US/09/300,985A
CURRENT FILING DATE: 1999-04-28
EARLIER APPLICATION NUMBER: US 08/955,937
EARLIER FILING DATE: 1997-10-22
EARLIER APPLICATION NUMBER: US 60/056,152
EARLIER FILING DATE: 1997-08-19
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 2345
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-300-985-1

Query Match 2.4%: Score 18; DB 4; Length 2345;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 GGCCCATCATTTGCCGAGA 428
|||||
DB 1986 GGCCCATCATTTGCCGAGA 1969

RESULT 9
US-09-020-956-32/c
Sequence 32, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillin, Davin C.

;; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;; NUMBER OF SEQUENCES: 178
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SEED AND BERRY LLP
;; STREET: 6300 Columbia Center, 701 Fifth Avenue
;; CITY: Seattle
;; STATE: WA
;; COUNTRY: USA
;; ZIP: 98104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/020,956
;; FILING DATE: 09-FEB-1998
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Maki, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.427C2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 32:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 789 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cdna
;; US-09-020-956-32

Query Match 2.3%; Score 17; DB 4; Length 789;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 AGCCTGTTGCCGTCCTG 514
|||||
DB 131 AGCCTGTTGCCGTCCTG 115

RESULT 10
US-09-030-607-32/c
; Sequence 32, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 32:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 789 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cdna
;; US-09-030-607-32

Query Match 2.3%; Score 17; DB 4; Length 789;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 AGCCTGTTGCCGTCCTG 514
|||||
DB 131 AGCCTGTTGCCGTCCTG 115

RESULT 11
US-09-439-313-32/c
; Sequence 32, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 789
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(789)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-32

Query Match 2.3%; Score 17; DB 4; Length 789;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 AGCCTGTTGCCGTCCTG 514
|||||
DB 131 AGCCTGTTGCCGTCCTG 115

RESULT 12
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 2.3% Score 17; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 TGTTCGCGCTCTGGGAGA 519
|||||
DB 1139255 tgttcgctctctgggaga 1139272

RESULT 13
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 2.3% Score 17; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 TGTTCGCGCTCTGGGAGA 519
|||||
DB 1139230 tgttcgctctctgggaga 1139247

RESULT 14
US-09-534-638-4/c
Sequence 4, Application US/095334638
Patent No. 6320038
GENERAL INFORMATION:
APPLICANT: Panula, Pertti A.J.
APPLICANT: Brandt, Annika
APPLICANT: Westerlund, Johanna
TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof
FILE REFERENCE: 2530-104
CURRENT APPLICATION NUMBER: US/09/534.638

CURRENT FILING DATE: 2000-03-27
EARLIER APPLICATION NUMBER: 09/365755
EARLIER FILING DATE: 1999-08-03
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 396
TYPE: DNA
ORGANISM: Mouse
FEATURE:
NAME/KEY: CDS
LOCATION: (7)...(393)
US-09-534-638-4

Query Match 2.1% Score 16; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 617 AGCACTGTATCTCTCC 632
|||||
DB 134 AGCACTGTATCTCTCC 119

RESULT 15
US-09-328-111-142
Sequence 142, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dettl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328.111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 142
LENGTH: 601
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(601)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-142

Query Match 2.1% Score 16; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TTCAGCTTTGGAACT 21
|||||
DB 243 ttacgctttggaaact 258

RESULT 16
US-09-227-357-87
Sequence 87, Application US/09227357

Wed Aug 7 05:46:16 2002

Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P2010P1
CURRENT APPLICATION NUMBER: US/09/227,357
CURRENT FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684
EARLIER FILING DATE: 1998-07-07
EARLIER APPLICATION NUMBER: 60/051,926
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,793
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12

NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 87
LENGTH: 735
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (376)
OTHER INFORMATION: n equals a.t.g. or c
US-09-227-357-87

Query Match 2.1%, Score 16; DB 4; Length 735;
Best Local Similarity 100.0%; Pred. No. 1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 TCACCTTCCTGCTGG 560
|||||
DB 284 tcacttctctgctgg 299
|||||

RESULT 17
US-08-844-188-42
Sequence 42, Application US/08844188
Patent No. 6127180
GENERAL INFORMATION:
APPLICANT: Narva, Kenneth E.
APPLICANT: Schnepf, H. Ernest
APPLICANT: Knuth, Mark
APPLICANT: Pollard, Michael R.
APPLICANT: Cardineau, Guy
APPLICANT: Schwab, George E.
TITLE OF INVENTION: Pesticidal Toxins
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,188
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/633,993
FILING DATE: 19-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-703C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1152 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-844-188-42

Query Match 2.1%, Score 16; DB 3; Length 1152;

Best Local Similarity 100.0%; Pred. No. le+02; Mismatches 0; Indels 0; Gaps 0;

Matches 16; Conservative 0;

Qy 571 CCAATCATCAACAAG 586

Db 1039 CCAATCATCAACAAG 1054

RESULT 18

US-08-233-389C-4/c
; Sequence 4, Application US/08233389C
; Patent No. 5639855
; GENERAL INFORMATION:
; APPLICANT: KITAMURA, Kazuo
; APPLICANT: KANGAWA, Kenji
; APPLICANT: MATSUO, Hisayuki
; APPLICANT: ETO, Tanenao
; TITLE OF INVENTION: ADRENOMEDULLIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: C/O FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10020

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,389C
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
; NAME: HALEY Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: SHGN-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA

FEATURE:
; NAME/KEY: CDS
; LOCATION: 148..711
US-08-233-389C-4

Query Match 2.1%; Score 16; DB 1; Length 1493;

Best Local Similarity 100.0%; Pred. No. le+02; Mismatches 0; Indels 0; Gaps 0;

Matches 16; Conservative 0;

Qy 45 CTGGGACCTCAGATC 60

Db 645 CTGGGACCTCAGATC 630

RESULT 19

US-08-801-863-4/c
; Sequence 4, Application US/08801863
; Patent No. 5830703
; GENERAL INFORMATION:
; APPLICANT: KITAMURA, Kazuo
; APPLICANT: KANGAWA, Kenji
; APPLICANT: MATSUO, Hisayuki
; APPLICANT: ETO, Tanenao

TITLE OF INVENTION: ADRENOMEDULLIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: C/O FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10020

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,863
; FILING DATE: CONCURRENTLY HERewith
; CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
; NAME: HALEY Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: SHGN-5 DIV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA

FEATURE:
; NAME/KEY: CDS
; LOCATION: 148..711
US-08-801-863-4

Query Match 2.1%; Score 16; DB 2; Length 1493;

Best Local Similarity 100.0%; Pred. No. le+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 45 CTGGGACCTCAGATC 60

Db 645 CTGGGACCTCAGATC 630

RESULT 20

US-08-486-596A-4/c

; Sequence 4, Application US/08486596A

; Patent No. 5837823

; GENERAL INFORMATION:

; APPLICANT: KITAMURA, Kazuo

; APPLICANT: KANGAWA, Kenji

; APPLICANT: MATSUO, Hisayuki

; APPLICANT: ETO, Tanenao

; TITLE OF INVENTION: ADRENOMEDULLIN

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: C/O FISH & NEAVE

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/486,596A

; FILING DATE: JUNE 7, 1995

```
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: HALEY JR., James F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: SHGN-5 DIV1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 596-9000
: TELEFAX: (212) 596-9090
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1493 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 148..711
: US-08-486-596A-4

Query Match 2.1%; Score 16; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 CTGGGACCTCAGAGTC 60
Db 645 CTGGGACCTCAGAGTC 630

RESULT 21
US-09-004-713-4/C
: Sequence 4, Application US/09004713
: Patent No. 5910416
: GENERAL INFORMATION:
: APPLICANT: KITAMURA, Kazuo
: APPLICANT: KANGAWA, Kenji
: APPLICANT: MATSUO, Hisayuki
: APPLICANT: ETO, Tanenao
: TITLE OF INVENTION: ADRENOMEDULLIN
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: C/O FISH & NEAVE
: STREET: 1251 Avenue of the Americas
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10020
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/004,713
: FILING DATE: JANUARY 7, 1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: HALEY JR., James F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: SHGN-5 DIV2 CON
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 596-9090
: TELEFAX: (212) 596-9090
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1493 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS

: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: HALEY JR., James F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: SHGN-5 DIV1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 596-9000
: TELEFAX: (212) 596-9090
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1493 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS

Query Match 2.1%; Score 16; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 CTGGGACCTCAGAGTC 60
Db 645 CTGGGACCTCAGAGTC 630

RESULT 22
US-09-059-522-2
: Sequence 2, Application US/09059522
: Patent No. 6001607
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Yue, Henry
: APPLICANT: Guegler, Karl J.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: HUMAN GROWTH-ASSOCIATED METHYLTRANSFERASE
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/059,522
: FILING DATE: HERewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Cerrone, Michael C
: REGISTRATION NUMBER: 39,132
: REFERENCE/DOCKET NUMBER: PF-0495 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2057 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: ADRETUT05
: CLONE: 2496002
: US-09-059-522-2

Query Match 2.1%; Score 16; DB 3; Length 2057;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 GCCAAGCAGAGGCC 236
Db 218 GCCAAGCAGAGGCC 233

RESULT 23
```

```
US-09-382-027-2
; Sequence 2, Application US/09382027
; Patent No. 6132714
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN GROWTH-ASSOCIATED METHYLTRANSFERASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/382.027
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/059,522
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0495 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2057 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: ADRETUT05
; CLONE: 2496002
;
US-09-382-027-2

Query Match          2.1%   Score 16; DB 3; Length 2057;
Best Local Similarity 100.0%; Pred. No. le+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  221 GCCAGCAGCAGGCC 236
      |||||
DB  218 GCCAAGCAGCAGGCC 233

RESULT 24
US-08-844-188-39
; Sequence 39, Application US/08844188
; Patent No. 6127150
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Knuth, Mark
; APPLICANT: Pollard, Michael R.
; APPLICANT: Cardineau, Guy
; APPLICANT: Schwab, George E.
; TITLE OF INVENTION: Pesticidal Toxins
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Salliwanchik, Lloyd & Salliwanchik
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/481,626
; FILING DATE: 07-JUN-1995

US-09-382-027-2
```


RESULT 28

US-09-275-925-6
; Sequence 6, Application US/09275925
; Patent No. 6080406
; GENERAL INFORMATION:
; APPLICANT: James W. Baumgartner
; APPLICANT: Donald C. Foster
; APPLICANT: Frank J. Grant
; APPLICANT: Cindy A. Sprecher
; TITLE OF INVENTION: HEMATOPOIETIC CYTOKINE RECEPTOR
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,925
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31,648
; REFERENCE/DOCKET NUMBER: 95-31
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2589 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 11..1882
US-09-275-925-6

Query Match 2.18; Score 16; DB 3; Length 2589;
Best Local Similarity 100.0%; Pred. No. 1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 625 ATCCCTCCTGCTGCAG 640

Db 2225 ATCCCTCCTGCTGCAG 2240

RESULT 29

US-09-442-100-7/c
; Sequence 7, Application US/09442100
; Patent No. 6359193
; GENERAL INFORMATION:
; APPLICANT: Xu, Tian
; APPLICANT: Tao, Wufan
; APPLICANT: Wang, Welyi
; APPLICANT: Zhang, Sheng
; APPLICANT: Yu, Wan
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATs
; TITLE OF INVENTION: GENES AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York

STATE: New York
COUNTRY: USA
ZIP: 10035-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,100
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/411,111
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6523-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3155 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2943
US-09-442-100-7

Query Match 2.18; Score 16; DB 4; Length 3155;
Best Local Similarity 100.0%; Pred. No. 1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 532 CCCAACCCAGCATCTCA 547

Db 2415 CCCAACCCAGCATCTCA 2400

RESULT 30

US-08-968-751-3/c
; Sequence 3, Application US/08968751
; Patent No. 5948643
; GENERAL INFORMATION:
; APPLICANT: Rubinfield, Bonnie
; APPLICANT: Polakis, Paul G.
; APPLICANT: Ligenfelter, Carol
; APPLICANT: Vuong, Terilyn T.
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,751
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

NAME: Giotta, Gregory
REGISTRATION NUMBER: 32,028
REFERENCE/DOCKET NUMBER: ONYX1024 GG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 262-8710
TELEFAX: (510) 222-9758
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3256 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 34..2541
US-08-968-751-3

Query Match 2.1% Score 16; DR 2; Length 3256;
Best Local Similarity 100.0%; Pred. No. 1.e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 TCGTGAACAATGTCAT 267
|||||

DB 2459 TCGTGAACAATGTCAT 2444

RESULT 31

US-08-306-691B-44/c
Sequence 44, Application US/08306691B
Patent No. 5734039
GENERAL INFORMATION:
APPLICANT: Calabretta, Bruno
APPLICANT: Skorski, Tomasz
TITLE OF INVENTION: ANTISENSE
TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
STREET: Two Penn Center, Suite 1800
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 kb
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306,691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 3565 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-306-691B-44

Query Match 2.1% Score 16; DB 1; Length 3565;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 320 TCCTCATCTCTGGATG 335
|||||

DB 1055 TCCTCATCTCTGGATG 1040

RESULT 32

PCT-US93-06251-20/G
Sequence 20, Application PC/TUS9306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DiGiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3565 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-20

Query Match 2.1% Score 16; DB 5; Length 3565;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 320 TCCTCATCTCTGGATG 335
|||||

DB 1055 TCCTCATCTCTGGATG 1040

RESULT 33

US-09-050-159-130/G
Sequence 130, Application US/09050159A
Patent No. 6197505
GENERAL INFORMATION:
APPLICANT: No. 6197505berg, Leif T
APPLICANT: Andersson, Maria K
APPLICANT: Linstrom, Per H
TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
TITLE OF INVENTION: COMPOSITIONS FOR USE THEREOF
FILE REFERENCE: 1248/15042
CURRENT APPLICATION NUMBER: US/09/050,159A

```
; CURRENT FILING DATE: 1998-03-27
; EARLIER APPLICATION NUMBER: 60/042,930
; EARLIER FILING DATE: 1987-04-03
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 130
; LENGTH: 4020
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Angiotensin I converting enzyme mRNA
US-09-050-159-130

Query Match          2.1%; Score 16; DB 4; Length 4020;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 231 AGCCCGAGGAGCTGA 246
      |||||||
Db 3765 AGCCCGAGGAGCTGA 3750

RESULT 34
US-09-162-484-18/c
; Sequence 18, Application US/09162484
; Patent No. 6248724
; GENERAL INFORMATION:
; APPLICANT: Phillips, M. Ian
; APPLICANT: Mohuczy, Dagmara
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE COMPOSITIONS TARGETED TO
; FILE REFERENCE: UFLA-087/UFLA087P
; CURRENT APPLICATION NUMBER: US/09/162,484
; CURRENT FILING DATE: 1998-09-25
; EARLIER APPLICATION NUMBER: 60/059,661
; EARLIER FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 4024
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-162-484-18

Query Match          2.1%; Score 16; DB 4; Length 4024;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 231 AGCCCGAGGAGCTGA 246
      |||||||
Db 3765 AGCCCGAGGAGCTGA 3750

RESULT 35
US-08-716-351A-1/c
; Sequence 1, Application US/08716351A
; Patent No. 6033905
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based
; TITLE OF INVENTION: Retroviral Vectors
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,351A
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03784
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian,, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 152#0-128-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8535 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..8535
; OTHER INFORMATION: /standard_name= "GalV SEATO Genome"
US-08-716-351A-1

Query Match          2.1%; Score 16; DB 3; Length 8535;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 AAAGGCATCCTTGAGA 170
      |||||||
Db 3379 AAAGGCATCCTTGAGA 3364

RESULT 36
US-09-534-638-1/c
; Sequence 1, Application US/09534638
; Patent No. 6320038
; GENERAL INFORMATION:
; APPLICANT: Panula, Pertti A.J.
; APPLICANT: Brandt, Annika
; APPLICANT: Westerlund, Johanna
; TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof
; TITLE OF INVENTION: For therapy and diagnosis
; FILE REFERENCE: 2530-104
; CURRENT APPLICATION NUMBER: US/09/534,638
; CURRENT FILING DATE: 2000-03-27
; EARLIER APPLICATION NUMBER: 09/365755
; EARLIER FILING DATE: 1999-08-03
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 9840
; TYPE: DNA
; ORGANISM: Mouse
US-09-534-638-1

Query Match          2.1%; Score 16; DB 4; Length 9840;
Best Local Similarity 100.0%; Pred. No. 1.1e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 617 AGCACTGTATCCTCCC 632
      |||||||
Db 5737 AGCACTGTATCCTCCC 5722

RESULT 37
US-09-338-907-183/c
; Sequence 183, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
```



```
? APPLICANT: Blumenfeld, Marta
? APPLICANT: Ilya, Chumakov
? APPLICANT: Bougueleret, Lydie
? TITLE OF INVENTION: PROSTATE CANCER GENE
? FILE REFERENCE: GENSET 18CP1C
? CURRENT APPLICATION NUMBER: US/09/338,907
? EARLIER FILING DATE: 1998-06-23
? EARLIER APPLICATION NUMBER: 08/996,306
? EARLIER FILING DATE: 1997-12-22
? EARLIER APPLICATION NUMBER: 60/099,658
? EARLIER FILING DATE: 1998-09-09
? EARLIER APPLICATION NUMBER: 09/218,207
? EARLIER FILING DATE: 1998-12-22
? NUMBER OF SEQ ID NOS: 578
? SOFTWARE: Patent.pm
? SEQ ID NO 183
? LENGTH: 37950
? TYPE: DNA
? ORGANISM: Mus musculus
? FEATURE:
? NAME/KEY: exon
? LOCATION: 5259..5328
? OTHER INFORMATION: exon2
? FEATURE:
? NAME/KEY: exon
? LOCATION: 12675..12791
? OTHER INFORMATION: exon3
? FEATURE:
? NAME/KEY: exon
? LOCATION: 14621..14710
? OTHER INFORMATION: exon4
? FEATURE:
? NAME/KEY: exon
? LOCATION: 19822..19912
? OTHER INFORMATION: exon5
? FEATURE:
? NAME/KEY: exon
? LOCATION: 21789..21950
? OTHER INFORMATION: exon6
? FEATURE:
? NAME/KEY: exon
? LOCATION: 23387..23510
? OTHER INFORMATION: exon7
? FEATURE:
? NAME/KEY: exon
? LOCATION: 25520..26016
? OTHER INFORMATION: exon8
? US-09-338-907-183
```

```
Query Match 2.18; Score 16; DB 4; Length 37950;
Best Local Similarity 100.0%; Pred. No. 1.le-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 331 GGATGTCAAAGCTGAA 346
|||||
DB 32060 GGATGTCAAAGCTGAA 32045
```

```
RESULT 38
US-09-218-207-183/c
? Sequence 183, Application US/09218207
? Patent No. 6346381
? GENERAL INFORMATION:
? APPLICANT: Cohen, Daniel
? APPLICANT: Blumenfeld, Marta
? APPLICANT: Ilya, Chumakov
? APPLICANT: Bougueleret, Lydie
? TITLE OF INVENTION: Prostate cancer gene
? FILE REFERENCE: GENSET.018CP1
? CURRENT APPLICATION NUMBER: US/09/218,207
? CURRENT FILING DATE: 1998-12-22
? EARLIER APPLICATION NUMBER: 08/996,306
```

```
? EARLIER FILING DATE: 1997-12-22
? EARLIER APPLICATION NUMBER: 60/099,658
? EARLIER FILING DATE: 1998-09-09
? NUMBER OF SEQ ID NOS: 578
? SOFTWARE: Patent.pm
? SEQ ID NO 183
? LENGTH: 37950
? TYPE: DNA
? ORGANISM: Mus musculus
? FEATURE:
? NAME/KEY: exon
? LOCATION: 5259..5328
? OTHER INFORMATION: exon2
? FEATURE:
? NAME/KEY: exon
? LOCATION: 12675..12791
? OTHER INFORMATION: exon3
? FEATURE:
? NAME/KEY: exon
? LOCATION: 14621..14710
? OTHER INFORMATION: exon4
? FEATURE:
? NAME/KEY: exon
? LOCATION: 19822..19912
? OTHER INFORMATION: exon5
? FEATURE:
? NAME/KEY: exon
? LOCATION: 21789..21950
? OTHER INFORMATION: exon6
? FEATURE:
? NAME/KEY: exon
? LOCATION: 23387..23510
? OTHER INFORMATION: exon7
? FEATURE:
? NAME/KEY: exon
? LOCATION: 25520..26016
? OTHER INFORMATION: exon8
? US-09-218-207-183
```

```
Query Match 2.18; Score 16; DB 4; Length 37950;
Best Local Similarity 100.0%; Pred. No. 1.le-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 331 GGATGTCAAAGCTGAA 346
|||||
DB 32060 GGATGTCAAAGCTGAA 32045
```

```
RESULT 39
US-09-103-840A-2/c
? Sequence 2, Application US/09103840A
? Patent No. 6294328
? GENERAL INFORMATION:
? APPLICANT: FLEISCHMAN, Robert D.
? APPLICANT: WHITE, Owen R.
? APPLICANT: FRASER, Claire M.
? APPLICANT: VENTER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? FILE REFERENCE: 24366-20007.00
? CURRENT APPLICATION NUMBER: US/09/103,840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: Patentin Ver. 2.1
? SEQ ID NO 2
? LENGTH: 4403765
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? FEATURE:
? OTHER INFORMATION: CDC 1551
? OTHER INFORMATION: "n" bases at various positions throughout the sequence
? OTHER INFORMATION: represent a, t, c or g
```

US-09-103-840A-2

Query Match 2.1%; Score 16; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 CTGAGAAATGCTGAA 258

Db 1708955 CTGAGAAATGCTGAA 1708940

RESULT 40

US-08-373-124A-128
; Sequence 128, Application US/08373124A
; Patent No. 5646042
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,124A
; FILING DATE: January 13, 1995

PRIOR APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992

ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-373-124A-128

Query Match 2.0%; Score 15; DB 1; Length 46;
Best Local Similarity 73.3%; Pred. No. 2.9e+02;
Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 206 AGTGCTTGGCAACTG 220

Db 1 AGUGCUUGGCAACUG 15

RESULT 41

US-08-435-628-128
; Sequence 128, Application US/08435628
; Patent No. 5817796
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,628
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 514

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,124
; FILING DATE: January 13, 1995
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992

ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-435-628-128

Query Match 2.0%; Score 15; DB 1; Length 46;
Best Local Similarity 73.3%; Pred. No. 2.9e+02;
Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 206 AGTGCTTGGCAACTG 220

Db 1 AGUGCUUGGCAACUG 15

RESULT 42

US-08-998-099-289
 ; Sequence 289, Application US/08998099A
 ; Patent No. 6103890
 ; GENERAL INFORMATION:
 ; APPLICANT: JARVIS, THALE
 ; APPLICANT: MCSWIGEN, JAMES A.
 ; APPLICANT: STINCHCOMB, DAN T.
 ; TITLE OF INVENTION: ENZYMATIC NUCLEIC ACID TREATMENT OF DISEASES
 ; FILE REFERENCE: 231/175
 ; CURRENT APPLICATION NUMBER: US/08/998,099A
 ; CURRENT FILING DATE: 1997-12-24
 ; EARLIER FILING DATE: 1997-01-23
 ; EARLIER APPLICATION NUMBER: 60/037,658
 ; EARLIER FILING DATE: 1995-01-13
 ; EARLIER APPLICATION NUMBER: 08/373,124
 ; EARLIER FILING DATE: 1994-05-18
 ; EARLIER APPLICATION NUMBER: 08/245,466
 ; NUMBER OF SEQ ID NOS: 375
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 289
 ; LENGTH: 50
 ; TYPE: RNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthesized Hairpin Ribozyme
 US-08-998-099-289

Query Match 2.0%; Score 15; DB 3; Length 50;
 Best Local Similarity 86.7%; Pred. No. 2.9e+02;
 Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 633 TGTGTCAGAGGAGA 647
 DB 1 UGUGCAGAGGAGA 15
 :||:|||||||

RESULT 43
 US-08-477-928A-43
 ; Sequence 43, Application US/08477928A
 ; Patent No. 6207389
 ; GENERAL INFORMATION:
 ; APPLICANT: Dosch, Hans M.
 ; TITLE OF INVENTION: METHODS FOR CONTROLLING T
 ; TITLE OF INVENTION: LYMPHOCYTE MEDIATED IMMUNE RESPONSES
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BAKER & BOTTS
 ; STREET: 1299 Pennsylvania Avenue
 ; CITY: Washington D.C.
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 20004-2400
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,928A
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Remick, James
 ; REGISTRATION NUMBER: 36902
 ; REFERENCE/DOCKET NUMBER: 19060-0105
 ; TELEPHONE: (202) 639 7700
 ; TELEFAX: (202) 639 7890
 ; INFORMATION FOR SEQ ID NO: 43:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 81 base pairs

; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 US-08-477-928A-43
 Query Match 2.0%; Score 15; DB 4; Length 81;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 571 CCAATCATCAACAA 585
 DB 46 CCAATCATCAACAA 60
 :|||||||

RESULT 44
 US-08-943-731-12
 ; Sequence 12, Application US/08943731
 ; Patent No. 6265157
 ; GENERAL INFORMATION:
 ; APPLICANT: PROCKOP, DARWIN J.
 ; APPLICANT: SPOTILA, LORETTA D.
 ; APPLICANT: DELTAS, CONSTANTINOS D.
 ; APPLICANT: SEREDA, LARISA
 ; APPLICANT: LARSON, ANDREA W.
 ; APPLICANT: PACK, MICHAEL
 ; APPLICANT: COLIGE, ALAIN
 ; APPLICANT: EARLY, JAMES
 ; APPLICANT: KORKKO, JARMO
 ; APPLICANT: ALA-KOKKO, LEENA, et al.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
 ; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
 ; NUMBER OF SEQUENCES: 666
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
 ; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
 ; STREET: FLR.
 ; CITY: PHILADELPHIA
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103-7086
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/943,731
 ; FILING DATE: 03-OCT-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/212,322
 ; FILING DATE: 14-MAR-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/803,628
 ; FILING DATE: 03-DEC-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DOYLE LEARY P.H.D., KATHRYN
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: 9598-27
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-965-1284
 ; TELEFAX: 215-567-2991
 ; TELEX: 831-494
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 227 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 US-08-943-731-12

Query Match 2.0%; Score 15; DB 4; Length 227;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 TGGGACCTCAGAGTC 60
|||||
DB 105 TGGGACCTCAGAGTC 119

RESULT 45

US-09-240-274-189
Sequence 189, Application US/09240274
Patent No. 6255455

GENERAL INFORMATION:

APPLICANT: Siegel, Donald L.
TITLE OF INVENTION: RH(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
FILE OF INVENTION: SORTING METHOD FOR PRODUCTION THEREOF
FILE REFERENCE: 09596-4202
CURRENT APPLICATION NUMBER: US/09/240.274
CURRENT FILING DATE: 1999-01-29
EARLIER APPLICATION NUMBER: 60/081,380
EARLIER FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/028,550
EARLIER FILING DATE: 1996-10-11
NUMBER OF SEQ ID NOS: 224
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 189
LENGTH: 378

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: anti-Rh(D) antibody clone SH28, SH50, and SH53

US-09-240-274-189

Query Match 2.0%; Score 15; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 TCTCCTGTGCGCGCT 39
|||||
DB 59 TCTCCTGTGCGCGCT 73

RESULT 46

US-08-557-309B-16
Sequence 16, Application US/08557309B
Patent No. 5916572

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodges, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435

PREVENTION OF T

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-557-309B-16

Query Match 2.0%; Score 15; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 633 TGCTGCAGAGGAGA 647
|||||
DB 3 TGCTGCAGAGGAGA 17

RESULT 47

US-08-834-306-16
Sequence 16, Application US/08834306
Patent No. 6054135

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodges, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:

LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-834-306-16

Query Match 2.0%; Score 15; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 633 TGCTGCAGAGGAGA 647

Db 3 TGCTGAGAGGAGGA 17
|||||

RESULT 48

US-08-993-674A-16
; Sequence 16, Application US/0893674A
; Patent No. 6228372

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

; NUMBER OF SEQUENCES: 81

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/993,674A

; FILING DATE: 18-DEC-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Naki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.422C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 456 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-993-674A-16

Query Match

Best Local Similarity 2.0%; Score 15; DB 4; Length 456;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 633 TGCTGAGAGGAGGA 647

Db 3 TGCTGAGAGGAGGA 17

RESULT 49

US-09-323-472A-5/c
; Sequence 5, Application US/09323472A
; Patent No. 6346382

GENERAL INFORMATION:

; APPLICANT: Summar, Marshall

; APPLICANT: Christman, Brian

; TITLE OF INVENTION: HUMAN CARBAMYL PHOSPHATE SYNTHETASE I POLYMORPHISM AND DIAGNOST

; FILE REFERENCE: 1242/19

; CURRENT APPLICATION NUMBER: US/09/323,472A

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO. 5

Query Match

Best Local Similarity 2.0%; Score 15; DB 1; Length 520;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 AACTCCCTCATCCTG 331

Db 5 AACTCCCTCATCCTG 19

Query Match

Best Local Similarity 2.0%; Score 15; DB 4; Length 495;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 AGAAATTGCTGAACA 260

Db 25 AGAAATTGCTGAACA 11

RESULT 50

US-07-743-518-25
; Sequence 25, Application US/07743518
; Patent No. 5397696

GENERAL INFORMATION:

; APPLICANT: YANAGIHARA, RICHARD

; APPLICANT: NERURKAR, VIVEK R.

; APPLICANT: JENKINS, CAROL

; APPLICANT: MILLER, MARK

; APPLICANT: GARRUTO, RALPH M.

; TITLE OF INVENTION: PAPUA NEW GUINEA HUMAN T-LYMPHOTROPIC

; TITLE OF INVENTION: VIRUS

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN, DAREY & CUSHMAN

; STREET: 1615 L STREET, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20036-5601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/743,518

; FILING DATE: 19910812

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: SCOTT, WATSON T.

; REGISTRATION NUMBER: 26,581

; REFERENCE/DOCKET NUMBER: WTS/5683/84699/SAP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-861-3000

; TELEFAX: 202-822-0944

; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 520 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-07-743-518-25

RESULT 51

US-08-721-488-7
 : Sequence 7, Application US/08721488
 : Patent No. 5965388
 : GENERAL INFORMATION:
 : APPLICANT: Jacobs, Kenneth
 : APPLICANT: McCoy, John
 : APPLICANT: LaValle, Edward
 : APPLICANT: Racie, Lisa
 : APPLICANT: Merberg, David
 : APPLICANT: Treacy, Maurice
 : APPLICANT: Spaulding, Vikki
 : APPLICANT: Bowman, Michael
 : TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 : TITLE OF INVENTION: ENCODING THEM
 : NUMBER OF SEQUENCES: 16
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Genetics Institute, Inc.
 : STREET: 87 Cambridgepark Drive
 : CITY: Cambridge
 : STATE: Massachusetts
 : COUNTRY: U.S.A.
 : ZIP: 02140
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/721,488
 : FILING DATE:
 : CLASSIFICATION: 424
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Brown, Scott A.
 : REGISTRATION NUMBER: 32,724
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617) 498-8224
 : TELEFAX: (617) 876-5851
 : INFORMATION FOR SEQ ID NO: 7:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 553 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : US-08-721-488-7

Query Match 2.0%; Score 15; DB 2; Length 553;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 AGAAGGCCGAGGAAG 242
 |||||
 Db 390 AGAAGGCCGAGGAAG 404

RESULT 52

US-08-803-899-1/c
 : Sequence 1, Application US/08803899
 : Patent No. 5912224
 : GENERAL INFORMATION:
 : APPLICANT: DONAHOE, PATRICIA K.
 : APPLICANT: WANG, TONGWEN
 : TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
 : TITLE OF INVENTION: CELLULAR RESPONSE TO TGF-BETA LIGANDS
 : NUMBER OF SEQUENCES: 26
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
 : STREET: 1100 NEW YORK AVENUE, SUITE 600
 : CITY: WASHINGTON

STATE: DC
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/803,899
 FILING DATE: 02/21/1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/012,054
 FILING DATE: 02/22/1996
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: STEFFE, ERIC K.
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 0609.4240001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)371-2600
 TELEFAX: (202)371-2540
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 571 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..324
 US-08-803-899-1

Query Match 2.0%; Score 15; DB 2; Length 571;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 GCAACTCCCTCATCC 329
 |||||
 Db 194 GCAACTCCCTCATCC 180

RESULT 53

US-09-385-982-231/c
 : Sequence 231, Application US/09385982
 : Patent No. 6262334
 : GENERAL INFORMATION:
 : APPLICANT: ENDEGE, WILSON O., ET AL.
 : TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 : TITLE OF INVENTION: PRODUCTS: II
 : FILE REFERENCE: CCDNA-260XX
 : CURRENT APPLICATION NUMBER: US/09/385,982
 : CURRENT FILING DATE: 1999-08-30
 : EARLIER APPLICATION NUMBER: 09/328,111
 : EARLIER FILING DATE: 1999-06-08
 : EARLIER APPLICATION NUMBER: 60/117,393
 : EARLIER FILING DATE: 1999-01-27
 : EARLIER APPLICATION NUMBER: 60/098,639
 : EARLIER FILING DATE: 1998-08-31
 : NUMBER OF SEQ ID NOS: 544
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 231
 : LENGTH: 601
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1)...(601)
 : OTHER INFORMATION: n = A,T,C or G
 : US-09-385-982-231

```
Query Match      2.0%; Score 15; DB 4; Length 601;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels
```

Qy 568 CAGCCAAATCATCA 582
Db 309 CAGCCAAATCATCA 295

```

RESULT 54
US-08-998-416-604/c
; Sequence 604, Application US/08998416
; Patent No. 6239284
; GENERAL INFORMATION:
; APPLICANT: Philippssen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Robischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; AND USES THEREOF

```

Query Match 2.0%; Score 15; DB 4; Length 728;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels

Qy	224	AAGCAGAAGGCCAG	238
D _b	89	AAGCAGAAGGCCAG	75

```

1  RESULT 55
2  US-08-173-510B-95
3  ; Sequence 95, Application US/08173510B
4  ; Patent No. 5747296
5  ; GENERAL INFORMATION:
6  ; APPLICANT: MATTHEW MOYLE, ET AL.
7  ; TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS
8  ; NUMBER OF SEQUENCES: 104
9  ; CORRESPONDENCE ADDRESS:
10 ; ADDRESSEE: Lyon & Lyon
11 ; STREET: 633 West Fifth Street
12 ; STREET: Suite 4700
13 ; CITY: Los Angeles
14 ; STATE: California
15 ; COUNTRY: U.S.A.
16 ; ZIP: 90071
17 ;
18 ; COMPUTER READABLE FORM:
19 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
20 ; MEDIUM TYPE: storage
21 ; COMPUTER: IBM Compatible
22 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
23 ; SOFTWARE: Word Perfect 5.1
24 ; CURRENT APPLICATION DATA:
25 ; APPLICATION NUMBER: US/08/173,510B
26 ; FILING DATE: 23-DEC-1993
27 ;
28 ; PRIOR APPLICATION DATA:
29 ; APPLICATION NUMBER: 08/151,064
30 ; FILING DATE: 10-NOV-1993
31 ; APPLICATION NUMBER: 08/060,433
32 ; FILING DATE: 11-MAY-1993
33 ; APPLICATION NUMBER: 07/996,972
34 ; FILING DATE: 24-DEC-1992
35 ; APPLICATION NUMBER: 07/881,721
36 ; FILING DATE: 11-MAY-1992
37 ; ATTORNEY/AGENT INFORMATION:
38 ; NAME: BIGGS, SUZANNE L.
39 ; REGISTRATION NUMBER: 30,158
40 ; REFERENCE/DOCKET NUMBER: 205/073
41 ; TELECOMMUNICATION INFORMATION:
42 ; TELEPHONE: (213) 489-1600
43 ; TELEFAX: (213) 955-0440
44 ;
45 ; INFORMATION FOR SEQ ID NO: 95:
46 ; SEQUENCE CHARACTERISTICS:
47 ; LENGTH: 877
48 ; TYPE: NUCLEIC ACID
49 ; STRANDEDNESS: SINGLE
50 ; TOPOLOGY: LINEAR
51 ; MOLECULE TYPE: NUCLEIC
52 ; FEATURE:
53 ; NAME/KEY: Coding Sequence
54 ; LOCATION: 1...822
55 ;
56 ; US-08-173-510B-95

```

Query Match 2.0%; Score 15; DB 1; Length 877;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels

Qy	345	AACGATCGATG	359
Dp	179	AACGATCGATG	193

RESULT 56
US-08-458-218-93
? Sequence 93, Application US/08458218
? Patent No. 5789178
? GENERAL INFORMATION:
? APPLICANT: MATTHEW MOYLE ET AL.
? TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS
? NUMBER OF SEQUENCES: 102
? CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,218
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/151,064
; FILING DATE: 10-NOVEMBER-1993
; APPLICATION NUMBER: 08/060,433
; FILING DATE: 11-MAY-1993
; APPLICATION NUMBER: 07/881,721
; FILING DATE: 11-MAY-1992
; APPLICATION NUMBER: 07/996,972
; FILING DATE: 24-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 203/226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 877
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: NUCLEIC
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...822
; US-08-458-218-93

Query Match 2.0%; Score 15; DB 1; Length 877;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 AACCGATCGATGATG 359
Db 179 AACCGATCGATGATG 193
|||||
RESULT 57
US-08-450-497-95
; Sequence 95, Application US/08/450497
; Patent No. 5919900
; GENERAL INFORMATION:
; APPLICANT: MATTHEW MOYLE, ET AL.
; TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,497
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/173,510
; FILING DATE: 23-DEC-1993
; APPLICATION NUMBER: 08/151,064
; FILING DATE: 10-NOV-1993
; APPLICATION NUMBER: 08/060,433
; FILING DATE: 11-MAY-1993
; APPLICATION NUMBER: 07/996,972
; FILING DATE: 24-DEC-1992
; APPLICATION NUMBER: 07/881,721
; FILING DATE: 11-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 205/073
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 95:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 877
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: NUCLEIC
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...822
; US-08-450-497-95

Query Match 2.0%; Score 15; DB 2; Length 877;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 345 AACCGATCGATGATG 359
Db 179 AACCGATCGATGATG 193
|||||
RESULT 58
US-09-446-504-4
; Sequence 4, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshiaki
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazuo
; APPLICANT: MUKAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1

```


SEQ ID NO 4
LENGTH: 984
TYPE: DNA
ORGANISM: Pyrococcus furiosus
US-09-446-504-4

Query Match
Best Local Similarity 100.0%; DB 4; Length 984;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 GCAGAGGAGATATG 651
Db 601 gcagaaggagatattg 615

RESULT 59

US-09-712-266-4
Sequence 4, Application US/09712266
Patent No. 633158
GENERAL INFORMATION:
APPLICANT: UEMORI, Takashi
APPLICANT: SATO, Yoshimi
APPLICANT: FUJITA, Tomoko
APPLICANT: MIYAKE, Kazuo
APPLICANT: MUKAI, Hiroyuki
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Kunoshin
TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
FILE REFERENCE: 1422-40BPCT
CURRENT APPLICATION NUMBER: US/09/712,266
CURRENT FILING DATE: 2000-11-15
PRIOR APPLICATION NUMBER: US 09/446,504
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: PCT/JP98/02845
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: JP 9-187496
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: JP 9-320692
PRIOR FILING DATE: 1997-11-27
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 984
TYPE: DNA
ORGANISM: Pyrococcus furiosus
US-09-712-266-4

Query Match
Best Local Similarity 100.0%; DB 4; Length 984;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 GCAGAGGAGATATG 651
Db 601 gcagaaggagatattg 615

RESULT 60

US-09-142-565-1
Sequence 1, Application US/09142565A
Patent No. 6187560
GENERAL INFORMATION:
APPLICANT: Lee James Beesley
APPLICANT: Kelly Paine
APPLICANT: Robert James
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GH-30002
CURRENT APPLICATION NUMBER: US/09/142,565A
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: 9704551.2
EARLIER FILING DATE: 1997-03-05
EARLIER APPLICATION NUMBER: 9705614.7

EARLIER FILING DATE: 1997-03-18
EARLIER APPLICATION NUMBER: 97305305.1
EARLIER FILING DATE: 1997-07-16
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1192
TYPE: DNA
ORGANISM: HOMO SAPIEN
US-09-142-565-1

Query Match
Best Local Similarity 100.0%; DB 4; Length 1192;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 CTCCTTGGAGCTCT 459
Db 146 CTCCTTGGAGCTCT 160

RESULT 61

US-08-579-667-3/c
Sequence 3, Application US/08579667
Patent No. 5705624
GENERAL INFORMATION:
APPLICANT: Fitzmaurice, Wayne P.
APPLICANT: Hellmann, Gary M.
APPLICANT: Grill, Laurence K.
APPLICANT: Kumagai, Monto H.
APPLICANT: Della-Cioppa, Guy R.
TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: 1211 East Morehead Street, PO Drawer 34009
CITY: Charlotte
STATE: NO. 5705624th Carolina
COUNTRY: USA
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,667
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 627-196
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1316 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1239
US-08-579-667-3

Query Match
Best Local Similarity 100.0%; DB 1; Length 1316;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 577 CATCAACAAGTTCGT 591
|||||
DB 490 CATCAACAAGTTCGT 476

RESULT 62

US-08-641-038A-1/c
; Sequence 1, Application US/08641038A
; Patent No. 5861154

GENERAL INFORMATION:

APPLICANT: SODA, KENJI
APPLICANT: TANAKA, HIDEHIKO
APPLICANT: INOUE, HIROYUKI
APPLICANT: INAGAKI, KENJI
APPLICANT: ESAKI, NOBUYOSHI

TITLE OF INVENTION: RECOMBINANT L-METHIONINE GAMMA-LYASE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

STREET: PO BOX 747

CITY: FALLS CHURCH

STATE: VA

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/641.038A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURPHY JR, GERALD M

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 0020-3989

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

FEATURE:

NAME/KEY: CDS

LOCATION: 61..1254

US-08-641-038A-1

Query Match 2.0%; Score 15; DB 2; Length 1320;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 CAGCAGGTCGTCGAT 712

|||||

DB 1221 CAGCAGGTCGTCGAT 1207

RESULT 63

US-09-059-178-1/c
; Sequence 1, Application US/09059178

; Patent No. 5863788

GENERAL INFORMATION:

APPLICANT: SODA, KENJI

APPLICANT: TANAKA, HIDEHIKO

APPLICANT: INOUE, HIROYUKI

APPLICANT: INAGAKI, KENJI

APPLICANT: ESAKI, NOBUYOSHI

TITLE OF INVENTION: RECOMBINANT L-METHIONINE GAMMA-LYASE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

STREET: PO BOX 747

CITY: FALLS CHURCH

STATE: VA

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURPHY JR, GERALD M

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 0020-3989

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

FEATURE:

NAME/KEY: CDS

LOCATION: 61..1254

US-08-641-038A-1

TITLE OF INVENTION: RECOMBINANT L-METHIONINE GAMMA-LYASE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

STREET: PO BOX 747

CITY: FALLS CHURCH

STATE: VA

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/059,178

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/641.038

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: MURPHY JR, GERALD M

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 0020-3989

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

FEATURE:

NAME/KEY: CDS

LOCATION: 61..1254

US-09-059-178-1

Query Match

2.0%; Score 15; DB 2; Length 1320;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 CAGCAGGTCGTCGAT 712

|||||

DB 1221 CAGCAGGTCGTCGAT 1207

RESULT 64

US-08-642-541-1/c

; Sequence 1, Application US/08642541

; Patent No. 5891704

GENERAL INFORMATION:

APPLICANT: YUYING, TAN

TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING

TITLE OF INVENTION: METHIONINASE, PURIFIED RECOMBINANT METHIONINASE PRODUCED

TITLE OF INVENTION: THEREFROM AND USES THEREOF

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 2000 Pennsylvania Avenue, NW

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/642,541
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Millman, Robert A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 31276-20002.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 48..1241
; US-08-642-541-1

Query Match 2.0%; Score 15; DB 2; Length 1369;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 CAGCAGGTCGTCGAT 712
DB 1208 CAGCAGGTCGTCGAT 1194
|||||

RESULT 65
US-08-642-541-3
; Sequence 3, Application US/08642541
; Patent No. 5891704
; GENERAL INFORMATION:
; APPLICANT: YUYING, TAN
; TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING
; TITLE OF INVENTION: METHIONINASE, PURIFIED RECOMBINANT METHIONINASE PRODUCED
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,541
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Millman, Robert A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 31276-20002.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: YES
; US-08-642-541-3

Query Match 2.0%; Score 15; DB 2; Length 1369;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 CAGCAGGTCGTCGAT 712
DB 162 CAGCAGGTCGTCGAT 176
|||||

RESULT 66
US-09-260-889-1/C
; Sequence 1, Application US/09260889
; Patent No. 6231854
; GENERAL INFORMATION:
; APPLICANT: TAN, YUYING
; TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING
; TITLE OF INVENTION: METHIONINASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,889
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/914,377
; FILING DATE: 19-AUG-1997
; APPLICATION NUMBER: 08/642,541
; FILING DATE: 03-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 31276-20002.11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 48..1241
; OTHER INFORMATION:
; US-09-260-889-1

Query Match 2.0%; Score 15; DB 4; Length 1369;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 CAGCAGGTCGTCGAT 712
|||||

```

Db 1208 CAGCAGGTCGTCCAT 1194

RESULT 67

US-09-260-889-3
; Sequence 3, Application US/09260889
; Patent No. 6231854
; GENERAL INFORMATION:
; APPLICANT: TAN, YUYING
; TITLE OF INVENTION: HIGH EXPRESSION MODULES ENCODING
; TITLE OF INVENTION: METHIONINASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260.889
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/914,377
; FILING DATE: 19-AUG-1997
; APPLICATION NUMBER: 08/642,541
; FILING DATE: 03-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 31276-20002.11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ANTI-SENSE: YES
US-09-260-889-3

Query Match 2.0%; Score 15; DB 4; Length 1369;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 698 CAGCAGGTCGTCCAT 712
Db 162 CAGCAGGTCGTCCAT 176

RESULT 68

US-08-174-467-20/c
; Sequence 20, Application US/08174467
; Patent No. 5451514
; GENERAL INFORMATION:
; APPLICANT: BOUDET, ALAIN M.
; APPLICANT: INZE, DIRK G.
; TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/174,467
; FILING DATE: 28-DEC-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/874,166
; FILING DATE: 27-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 95563/PS36321/US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-174-467-20

Query Match 2.0%; Score 15; DB 1; Length 1377;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 GAATGTCACGTGGC 408
Db 259 GAATGTCACGTGGC 245

RESULT 69

US-08-452-071-20/c
; Sequence 20, Application US/08452071
; Patent No. 6066780
; GENERAL INFORMATION:
; APPLICANT: BOUDET, ALAIN M.
; APPLICANT: INZE, DIRK G.
; TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,071
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/874,166

FILING DATE: 27-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 95563/PS36321/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-452-071-20

Query Match 2.0%; Score 15; DB 3; Length 1377;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 GAATGTCACGTGGC 408
|||||
DB 259 GAATGTCACGTGGC 245

RESULT 70
US-08-961-083-157
Sequence 157, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 157:
SEQUENCE CHARACTERISTICS:
LENGTH: 1415 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-157

Query Match 2.0%; Score 15; DB 3; Length 1415;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 573 AAATCATCAACAAGT 587
|||||
DB 1165 AAATCATCAACAAGT 1179

RESULT 71
US-07-829-954-1/C
Sequence 1, Application US/07829954
Patent No. 534819
GENERAL INFORMATION:
APPLICANT: HAMMOND, Geoffrey L.
TITLE OF INVENTION: Pharmaceutical Composition For Use In
Treating Inflammation
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/829,954
FILING DATE: 19920203
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/664,114
FILING DATE: 04-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/204,356
FILING DATE: 09-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/653,736
FILING DATE: 11-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/204,400
FILING DATE: 09-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/160
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1423 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1253
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 36..101
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 102..1253
US-07-829-954-1

Query Match 2.0%; Score 15; DB 1; Length 1423;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 GTCATTTCTAAGCTG 277

Db 412 GTCATTCTAAGCTG 398
|||||

RESULT 72

US-07-994-423-1/c
; Sequence 1, Application US/07994423
; Patent No. 5432080
; GENERAL INFORMATION:
; APPLICANT: HAMMOND, Geoffrey L.
; APPLICANT: AVAKUMOV, George V.
; TITLE OF INVENTION: Variants of Corticosteroid Binding
; TITLE OF INVENTION: Globulin
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 16-DEC-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/197/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1253
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 36..101
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 102..1253
US-07-994-423-1

Query Match 2.0%; Score 15; DB 1; Length 1423;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 GTCATTCTAAGCTG 277
|||||

Db 412 GTCATTCTAAGCTG 398

RESULT 73

US-08-421-891-1/c
; Sequence 1, Application US/08421891
; Patent No. 5595969
; GENERAL INFORMATION:
; APPLICANT: HAMMOND, Geoffrey L.
; APPLICANT: AVAKUMOV, George V.
; TITLE OF INVENTION: Variants of Corticosteroid Binding
; TITLE OF INVENTION: Globulin

; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 14-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/994,423
; FILING DATE: 16-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/197/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1253
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 36..101
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 102..1253
US-08-421-891-1

Query Match 2.0%; Score 15; DB 1; Length 1423;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 GTCATTCTAAGCTG 277
|||||

Db 412 GTCATTCTAAGCTG 398

RESULT 74

US-09-040-799-2/c
; Sequence 2, Application US/09040799
; Patent No. 5885820
; GENERAL INFORMATION:
; APPLICANT: CHANG, YIE-HWA
; TITLE OF INVENTION: CLONE OF A NUCLEOTIDE SEQUENCE ENCODING
; TITLE OF INVENTION: A PROTEIN HAVING TWO FUNCTIONS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,799
FILING DATE: 18-MAR-1998
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 16153-4639
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-040-799-2

Query Match 2.0%; Score 15; DB 2; Length 1434;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 630 CCTGCTGCACAGG 644
Db 156 CCTGCTGCACAGG 142
|||||

RESULT 75
US-09-444-053-3/c
Sequence 3, Application US/09444053A
Patent No. 6165728
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
TITLE OF INVENTION: ANTISENSE MODULATION OF NCK-2 EXPRESSION
FILE REFERENCE: RTS-0122
CURRENT APPLICATION NUMBER: US/09/444,053A
CURRENT FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 89
SEQ ID NO 3
LENGTH: 1457
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (133)..(1275)
US-09-444-053-3

Query Match 2.0%; Score 15; DB 4; Length 1457;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 GTTCTTCACGAGGA 130
Db 342 GTTCTTCACGAGGA 328
|||||

Search completed: August 6, 2002, 20:45:06
Job time: 8899 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:42:48 : Search time 19.11 seconds
(without alignments)
1252.027 Million cell updates/sec

Title: US-10-020-139-2

Perfect score: 249

Sequence: 1 MLQLMKLVLLCGVLTGTSES.....NVIIQVVDPNQKHTQLQLI 249

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database : PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DR	ID	Description
1	8	3.2	432	2	T39775	hypothetical prote
2	8	3.2	562	1	A41801	phosphoglucumutase
3	8	3.2	562	1	PMRB	phosphoglucumutase
4	8	3.2	562	1	PMRT	phosphoglucumutase
5	8	3.2	566	1	PMRI	phosphoglucumutase
6	8	3.2	604	2	T20875	hypothetical prote
7	8	3.2	706	2	F82148	hypothetical prote
8	8	3.2	817	2	A82511	glycogen phosphory
9	8	3.2	895	2	T49010	hypothetical prote
10	7	2.8	58	2	A2103	hypothetical prote
11	7	2.8	141	2	F11196	hypothetical prote
12	7	2.8	161	2	A2262	aliphococyanin B
13	7	2.8	163	1	S01399	H+-transporting tw
14	7	2.8	179	2	T06962	conserved hypothet
15	7	2.8	187	2	AH0877	hypothetical prote
16	7	2.8	206	2	H97465	hypothetical prote
17	7	2.8	206	2	A82684	acetyltransferase
18	7	2.8	207	2	G84983	5-amino-6-(5-phosp
19	7	2.8	211	1	C65080	hypothetical prote
20	7	2.8	211	2	H91106	hypothetical prote
21	7	2.8	211	2	C85952	hypothetical prote
22	7	2.8	213	2	A82785	hypothetical prote
23	7	2.8	217	2	A97565	hypothetical prote
24	7	2.8	235	2	B42337	parotid secretory
25	7	2.8	257	2	H97216	probable S-adenosy
26	7	2.8	262	2	A81504	conserved hypothet
27	7	2.8	262	2	B72037	ct598 hypothetical
28	7	2.8	262	2	E86588	ct598 hypothetical
29	7	2.8	267	2	S48317	hypothetical prote

30	7	2.8	270	2	AB0514	DnaJ-like protein
31	7	2.8	271	2	G64726	yabH protein - Esc
32	7	2.8	271	2	D90636	probable DNA bindi
33	7	2.8	271	2	D85487	probable DNA bindi
34	7	2.8	277	2	AF0061	DnaJ-like protein
35	7	2.8	290	2	T27690	hypothetical prote
36	7	2.8	305	2	T03558	ribose transport A
37	7	2.8	307	2	S59153	NADH dehydrogenase
38	7	2.8	315	2	AE2384	hypothetical prote
39	7	2.8	322	1	G70037	conserved hypothet
40	7	2.8	323	2	AH1383	conserved hypothet
41	7	2.8	343	2	AC1920	ABC phosphate tran
42	7	2.8	348	2	T13817	NADH dehydrogenase
43	7	2.8	374	2	F95861	hypothetical ABC t
44	7	2.8	381	2	S18543	hypothetical prote
45	7	2.8	383	2	F72352	hypothetical prote
46	7	2.8	389	2	D84068	hypothetical prote
47	7	2.8	390	2	B55889	isoaspartyl dipept
48	7	2.8	390	2	F91289	isoaspartyl dipept
49	7	2.8	390	2	C86131	isoaspartyl dipept
50	7	2.8	409	2	C64553	chlorohydrolase -
51	7	2.8	409	2	F71956	hypothetical prote
52	7	2.8	409	2	G81422	hypothetical prote
53	7	2.8	413	2	T21644	hypothetical prote
54	7	2.8	417	2	AG1934	folypolyglutamate
55	7	2.8	424	2	B81740	3,4-dihydroxy-2-bu
56	7	2.8	435	2	D71236	hypothetical prote
57	7	2.8	444	2	T04812	hypothetical prote
58	7	2.8	448	2	B64503	hypothetical prote
59	7	2.8	455	2	T38275	hypothetical prote
60	7	2.8	458	2	E82175	conserved hypothet
61	7	2.8	466	2	T41076	hypothetical prote
62	7	2.8	486	2	H84805	hypothetical prote
63	7	2.8	536	2	B91174	hypothetical prote
64	7	2.8	557	2	D98195	hypothetical prote
65	7	2.8	557	2	AF3091	hypothetical prote
66	7	2.8	557	2	B86020	hypothetical prote
67	7	2.8	558	2	G85253	hypothetical prote
68	7	2.8	588	1	ZFECP3	penicillin-binding
69	7	2.8	588	2	H90639	septicum formation p
70	7	2.8	588	2	H85490	penicillin-binding
71	7	2.8	588	2	AG0517	penicillin-binding
72	7	2.8	599	2	T39535	MFS efflux transpo
73	7	2.8	625	2	H98148	hypothetical prote
74	7	2.8	627	2	T47968	dynammin-like prote
75	7	2.8	632	2	T02627	hypothetical prote

ALIGNMENTS

RESULT 1

T39775
hypothetical protein SPBC18H10.11c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C:Accession: T39775
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z21879
A:Accession: T39775
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-432 <LYN>
A:Cross-references: EMBL:AL022304; PIDN:CAA18408.1; GSPDB:GN00067; SPDB:SPBC18H10.11c
A:Experimental source: strain 972h; cosmid c18H10
C:Genetics:
A:Gene: SPDB:SPBC18H10.11c
A:Map position: 2
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC18H10.11c

Query Match 3.2% Score 8; DB 2; Length 432;

Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 VSSLLQKE 215
| | | | | | | |
Db 84 VSSLLQKE 91
| | | | | | | |

RESULT 2

A41801
phosphoglucomutase (EC 5.4.2.2) 1 - human
N:Alternate names: PGM1
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 19-May-1994 #text_change 18-Jun-1999
C:Accession: A41801
R:Whitehouse, D.B.; Putt, W.; Lovegrove, J.U.; Morrison, K.; Hollyoake, M.; Fox, M.F.; H
Proc. Natl. Acad. Sci. U.S.A. 89, 411-415, 1992
A:Title: Phosphoglucomutase 1: complete human and rabbit mRNA sequences and direct mappi
A:Reference number: A41801; MUID:92108065
A:Accession: A41801
A:Molecule type: mRNA
A:Residues: 1-562 <WHI>
A:Cross-references: GB:M83088; NID:g189925; PIDN:AAA60080.1; PID:g189926
A:Note: sequence extracted from NCBI backbone (NCBI:75120, NCBI:P75121)
C:Comment: This enzyme interconverts glucose-1-phosphate and glucose-6-phosphate, partic
C:Comment: This protein is a major substrate for Ca⁺⁺/calmodulin-dependent protein kinas
C:Genetics:
A:Gene: GDB:PGM1
A:Cross-references: GDB:119489; OMIM:171900
A:Map position: 1p31-1p31
A:Note: locus is highly polymorphic
C:Superfamily: phosphoglucomutase
C:Keywords: intramolecular transferase; isomerase; magnesium; monomer; phosphoprotein; S
F:2-562/Product: phosphoglucomutase 1 #status predicted <MAT>
F:117/Active site: Ser (phosphoserine intermediate) #status predicted
F:185/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status
F:408/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status
F:460/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status
F:467/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status

Query Match 3.2%; Score 8; DB 1; Length 562;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 LKVDLGVL 65
| | | | | | | |
Db 163 LKVDLGVL 170
| | | | | | | |

RESULT 3

PMRB
phosphoglucomutase (EC 5.4.2.2) 1, short splice form - rabbit
N:Alternate names: PGM1; sarcoplasmic reticulum-associated phosphoglucomutase; SR-PGM
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 03-Aug-1984 #sequence_revision 19-May-1994 #text_change 18-Jun-1999
C:Accession: B41801; B45077; A01175; B60810
R:Whitehouse, D.B.; Putt, W.; Lovegrove, J.U.; Morrison, K.; Hollyoake, M.; Fox, M.F.; H
Proc. Natl. Acad. Sci. U.S.A. 89, 411-415, 1992
A:Title: Phosphoglucomutase 1: complete human and rabbit mRNA sequences and direct mappi
A:Reference number: A41801; MUID:92108065
A:Accession: B41801
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-562 <WHI>
R:Lee, Y.S.; Marks, A.R.; Gureckas, N.; Lacro, R.; Nadali-Ginard, B.; Kim, D.H.
J. Biol. Chem. 267, 21080-21088, 1992
A:Title: Purification, characterization, and molecular cloning of a 60-kDa phosphoprotei
A:Reference number: A45077; MUID:93016027
A:Accession: B45077
A:Molecule type: mRNA
A:Residues: 1-562 <LEE>
A:Cross-references: GB:M97664; NID:g165663; PIDN:AAA31454.1; PID:g165664

A:Experimental source: skeletal muscle
A:Note: authors translated the codon GAG for residue 436 as Gln and CAG for residue 5
A:Note: these authors refer to this form as isoform 2 or PGM2
R:Ray Jr., W.J.; Hermodson, M.A.; Puvathingal, J.M.; Mahoney, W.C.
J. Biol. Chem. 258, 9166-9174, 1983
A:Title: The complete amino acid sequence of rabbit muscle phosphoglucomutase.
A:Reference number: A92435; MUID:83265705
A:Accession: A01175
A:Molecule type: protein
A:Residues: 2-562 <RAY>
R:Yuen, S.; Hunkapiller, M.W.; Wilson, K.J.; Yuan, P.M.
Anal. Biochem. 168, 5-15, 1988
A:Title: Applications of tandem microprobe liquid chromatography and sodium dodecyl su
A:Reference number: A60810; MUID:88207910
A:Accession: B60810
A:Molecule type: protein
A:Residues: 2-21 <YUE>
R:Milstein, C.P.; Milstein, C.
Biochem. J. 109, 93-99, 1968
A:Title: A tryptic peptide containing a unique serine phosphate residue in rabbit pho
A:Reference number: A90241; MUID:68368401
A:Contents: annotation; active site
C:Comment: This enzyme interconverts glucose-1-phosphate and glucose-6-phosphate, par
C:Comment: This protein is a major substrate for Ca⁺⁺/calmodulin-dependent protein ki
C:Superfamily: phosphoglucomutase
C:Keywords: alternative splicing; intramolecular transferase; isomerase; magnesium; m
F:2-562/Product: phosphoglucomutase 1, short splice form #status experimental
F:117/Active site: Ser (phosphoserine intermediate) #status experimental
F:185/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #stat
F:408/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #stat
F:460/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #stat
F:467/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #stat

Query Match 3.2%; Score 8; DB 1; Length 562;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 LKVDLGVL 65
| | | | | | | |
Db 163 LKVDLGVL 170
| | | | | | | |

RESULT 4

PMRT
phosphoglucomutase (EC 5.4.2.2) 1 - rat
N:Alternate names: PGM1
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-May-1994 #sequence_revision 02-Jun-1994 #text_change 18-Jun-1999
C:Accession: JC2011
R:Rivera, A.A.; Elton, T.S.; Dey, N.B.; Rounellis, P.; Marchase, R.R.
Gene 133, 261-266, 1993
A:Title: Isolation and expression of a rat liver cDNA encoding phosphoglucomutase.
A:Reference number: JC2011; MUID:94040821
A:Accession: JC2011
A:Molecule type: mRNA
A:Residues: 1-562 <RIV>
A:Cross-references: GB:L11694; NID:g393212; PIDN:AAA16862.1; PID:g393213
A:Experimental source: liver
A:Note: the authors translated the codon GAC for residue 62 as Cys
C:Comment: This enzyme interconverts glucose-1-phosphate and glucose-6-phosphate, par
C:Comment: This protein is a major substrate for Ca⁺⁺/calmodulin-dependent protein ki
C:Superfamily: phosphoglucomutase
C:Keywords: intramolecular transferase; isomerase; magnesium; monomer; phosphoprotein
F:2-562/Product: phosphoglucomutase 1 #status predicted <MAT>
F:117/Active site: Ser (phosphoserine intermediate) #status predicted
F:185/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #stat
F:460/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #stat
F:467/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #stat

Query Match 3.2%; Score 8; DB 1; Length 562;
Best Local Similarity 100.0%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LKVDLGVL 65
 |||||
 Db 163 LKVDLGVL 170

RESULT 5
 PMRBI
 phosphoglucomutase (EC 5.4.2.2) 1, long splice form - rabbit
 N:Alternate names: PGM
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 17-Feb-1994 #sequence_revision 02-Jun-1994 #text_change 18-Jun-1999
 C:Accession: AF5077
 R:Lee, Y.S.; Marks, A.R.; Gureckas, N.; Lacro, R.; Nadal-Ginard, B.; Kim, D.H.
 J. Biol. Chem. 267, 21080-21088, 1992
 A:Title: Purification, characterization, and molecular cloning of a 60-kDa phosphoprotein
 A:Reference number: A45077; MUID:99016027
 A:Accession: A45077
 A:Molecule type: mRNA
 A:Residues: 1-566 <LRE>
 A:Cross-references: GB:M97663; NID:g165661; PIDN:AAA31453.1; PID:g165662
 A:Experimental source: skeletal muscle
 A:Note: authors translated the codon CAG for residue 440 as Gln and CAG for residue 534
 A:Note: these authors refer to this form as isoform 1 or PGM1
 C:Comment: This enzyme interconverts glucose-1-phosphate and glucose-6-phosphate, partic
 C:Superfamily: phosphoglucomutase
 C:Keywords: alternative splicing; intramolecular transferase; isomerase; magnesium; mon
 F:6-566/Product: phosphoglucomutase 1, long splice form #status predicted <KAT>
 F:121/Active site: Ser (phosphoserine intermediate) #status predicted
 F:189/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status
 F:412/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status
 F:464/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status
 F:471/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status

Query Match 3.2%; Score 8; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LKVDLGVL 65
 |||||
 Db 167 LKVDLGVL 174

RESULT 6
 T20875
 hypothetical protein F14B6.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20875
 R:White, S.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19338
 A:Accession: T20875
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-604 <WIL>
 A:Cross-references: EMBL:Z81502; PIDN:CAB04106.1; GSPDB:GN00019; CESP:F14B6.2
 A:Experimental source: clone F14B6
 C:Genetics:
 A:Gene: CESP:F14B6.2
 A:Map position: 1
 A:Introns: 32/2; 93/2; 205/2; 469/3

Query Match 3.2%; Score 8; DB 2; Length 604;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 EAEKLLNN 87
 |||||
 Db 285 EAEKLLNN 292

RESULT 7
 F82148
 methyl-accepting chemotaxis protein VCI1859 [imported] - Vibrio cholerae (strain N169)
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: F82148
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
 I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833
 A:Accession: F82148
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-706 <HEI>
 A:Cross-references: GB:AE004261; GB:AE003852; NID:g9656382; PIDN:AAF95007.1; GSPDB:GI
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VCI1859
 A:Map position: 1

Query Match 3.2%; Score 8; DB 2; Length 706;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 STVSSLLQ 213
 |||||
 Db 286 STVSSLLQ 293

RESULT 8
 A82511
 glycogen phosphorylase (EC 2.4.1.1) [similarity] - Vibrio cholerae (strain N16961 se;
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 24-Aug-2001
 C:Accession: A82511
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
 I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833
 A:Accession: A82511
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-817 <HEI>
 A:Cross-references: GB:AE004345; GB:AE003853; NID:g9657390; PIDN:AAF95927.1; GSPDB:GI
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VCA0013
 A:Map position: 2
 A:Superfamily: phosphorylase
 C:Keywords: glycosyltransferase; hexosyltransferase; phosphoprotein; pyridoxal phosph
 F:664/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 3.2%; Score 8; DB 2; Length 817;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 LKASLDLL 153
 |||||
 Db 722 LKASLDLL 729

RESULT 9
 T49010
 hypothetical protein F25L23.280 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
 C:Accession: T49010
 R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
 Submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25012
 A:Accession: T49010
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-895 <DAN>
 A:Cross-references: EMBL:AL356014; GSPDB:GN00061; ATSP:F25L23.280
 A:Experimental source: cultivar Columbia; BAC clone F25L23
 C:Genetics:
 A:Gene: ATSP:F25L23.280
 A:Map position: 3
 C:Superfamily: S-receptor kinase; protein kinase homology; S-locus-specific glycoprotein

Query Match 3.2% Score 8; DB 2; Length 895;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 CGVLTGTS 18
 |||||
 Db 312 CGVLTGTS 319

RESULT 10
 AH2103
 hypothetical protein asl2383 [imported] - Anabaena sp. (strain PCC 7120)
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AH2103
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriugu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2103
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-58 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA74082.1; PID:gl7131475; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: asl2383

Query Match 2.8% Score 7; DB 2; Length 58;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 78 AQBAEKL 84
 |||||
 Db 19 AQBAEKL 25

RESULT 11
 C71196
 hypothetical protein PH1841 - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
 C:Accession: C71196
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohtsuka, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
 DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
 A:Reference number: A71000; MUID:98344137
 A:Accession: C71196
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-141 <KAW>
 A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA30962.1; PID:g3258279

A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
 C:Genetics:
 A:Gene: PH1841

Query Match 2.8% Score 7; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 178 PTSISLS 184
 |||||
 Db 51 PTSISLS 57

RESULT 12
 AF2262
 allophycocyanin B alpha chain [imported] - Anabaena sp. (strain PCC 7120)
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AF2262
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AF2262
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-161 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA75352.1; PID:gl7132786; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: apcd
 C:Superfamily: phycocyanin

Query Match 2.8% Score 7; DB 2; Length 161;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 KASLDLL 153
 |||||
 Db 135 KASLDLL 141

RESULT 13
 S01399
 H+-transporting two-sector ATPase (Ec 3.6.3.14) chain b precursor - thermophilic bact
 C:Species: thermophilic bacterium PS-3
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 14-Dec-2001
 C:Accession: S01399
 R:Ohta, S.; Yoshida, M.; Ishizuka, M.; Hirata, H.; Hamamoto, T.; Otawara-Hamamoto, Y.;
 Biochim. Biophys. Acta 933, 141-155, 1998
 A:Title: Sequence and over-expression of subunits of adenosine triphosphate synthase
 A:Reference number: S01397; MUID:88163679
 A:Accession: S01399
 A:Molecule type: DNA
 A:Residues: 1-163 <OHT>
 A:Cross-references: EMBL:X07804; NID:g45808; PIDN:CAA30650.1; PID:g581473
 A:Note: part of this sequence, including the amino end of the mature protein, was con
 C:Genetics:
 A:Start codon: GTG
 C:Superfamily: H+-transporting ATP synthase chain I
 C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; transmembrane p
 F:12-163/Product: H+-transporting ATP synthase chain b #status experimental <MAT>

Query Match 2.8% Score 7; DB 1; Length 163;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 QEAELK 85
|||||
DB 67 QEAELK 73

RESULT 14

T06962
hypothetical protein - Cyanophora paradoxa cyanelle
C:Species: cyanelle Cyanophora paradoxa
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 08-Oct-1999
R:Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bolnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A:Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A:Reference number: Z15840
A:Accession: T06962
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-179 <STJ>
A:Cross-references: EMBL:U30821; NID:gl016083; PIDN:AAA81305.1; PID:gi016218
A:Experimental source: strain Pringsheim LB555
C:Genetics:
A:Genome: cyanelle
C:Keywords: cyanelle

Query Match 2.8%; Score 7; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 HOPVAVL 171
|||||
DB 7 HOPVAVL 13

RESULT 15

AH0877
conserved hypothetical protein STV3249 [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH0877
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 846-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AH0502; PMID:11677608
A:Accession: AH0877
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-187 <PAK>
A:Cross-references: GB:AL513382; PIDN:CAD02920.1; PID:gl6504173; GSPDB:GN00176
C:Genetics:
A:Gene: STV3249
C:Superfamily: hypothetical protein H10304

Query Match 2.8%; Score 7; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 GILEKX 59
|||||
DB 50 GILEKX 56

RESULT 16

H97455
hypothetical protein AGR_C_1598 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C:Accession: H97465
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: H97465
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86681.1; PID:gl5155865; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_1598
A:Map position: circular chromosome

Query Match 2.8%; Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 STVSSLL 212
|||||
DB 30 STVSSLL 36

RESULT 17

AB2684
acetyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AB2684
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woc
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCI
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB2684
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAI41888.1; PID:gl7739251; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu0874
A:Map position: circular chromosome

Query Match 2.8%; Score 7; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 STVSSLL 212
|||||
DB 30 STVSSLL 36

RESULT 18

G84983
N-6-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193) [imported] - Buchner
A:Alternate names: riboflavin reductase
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: G84983
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp
A:Reference number: A84930; PMID:20445173
A:Accession: G84983
A:Status: preliminary
A:Molecule type: DNA

```

A:Residues: 1-207 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: RibD2; BU462
C:Keywords: oxidoreductase

Query Match          2.8%; Score 7; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LKISNSL 108
Db 178 LKISNSL 184
|||||

RESULT 19
C65080
hypothetical protein b2948 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: C65080
R:Blattner, F.K.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C65080
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <BLAT>
A:Cross-references: GB:AE000377; GB:U00096; NID:q2367178; PIDN:AAC75985.1; PID:g1789317;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: hypothetical protein HI0304

Query Match          2.8%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 GILEKLK 59
Db 74 GILEKLK 80
|||||

RESULT 20
H91106
hypothetical protein ECs3824 [imported] - Escherichia coli (strain O157:H7, substrain R)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: H91106
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H91106
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA037247.1; PID:g13363296; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs3824
C:Superfamily: hypothetical protein HI0304

Query Match          2.8%; Score 7; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 GILEKLK 59
Db 74 GILEKLK 80
|||||

A:Residues: 1-207 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: RibD2; BU462
C:Keywords: oxidoreductase

Query Match          2.8%; Score 7; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LKISNSL 108
Db 178 LKISNSL 184
|||||

RESULT 19
C65080
hypothetical protein b2948 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: C65080
R:Blattner, F.K.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C65080
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <BLAT>
A:Cross-references: GB:AE000377; GB:U00096; NID:q2367178; PIDN:AAC75985.1; PID:g1789317;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: hypothetical protein HI0304

Query Match          2.8%; Score 7; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 GILEKLK 59
Db 74 GILEKLK 80
|||||

RESULT 20
H91106
hypothetical protein ECs3824 [imported] - Escherichia coli (strain O157:H7, substrain R)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: H91106
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H91106
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA037247.1; PID:g13363296; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs3824
C:Superfamily: hypothetical protein HI0304

Query Match          2.8%; Score 7; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 GILEKLK 59
Db 74 GILEKLK 80
|||||

A:Residues: 1-211 <STO>
A:Cross-references: GB:AE005174; NID:g12517494; PIDN:AAG58079.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yggE
C:Superfamily: hypothetical protein HI0304

Query Match          2.8%; Score 7; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 GILEKLK 59
Db 74 GILEKLK 80
|||||

RESULT 22
AG2785
hypothetical protein pcm [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AG2785
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; MCCI
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG2785
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42701.1; PID:g17740137; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: pcm
A:Map position: circular chromosome

Query Match          2.8%; Score 7; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 LDLLTAV 156
Db 22 LDLLTAV 28
|||||

RESULT 23
AG97565
hypothetical protein AGR_C_3127 [imported] - Agrobacterium tumefaciens (strain C58, C
C:Species: Agrobacterium tumefaciens

```

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: A97565
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Alanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, D.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*
 A:Reference number: A97359; PMID:11743194
 A:Accession: A97565
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-217 <KUR>
 A:Cross-references: CH:AE007869; PIDN:AAK87474.1; PID:g15156796; GSPDB:GNO0169
 C:Genetics:
 A:Gene: AGR_C_3127
 A:Map position: circular chromosome

Query Match 2.8% Score 7; DB 2; Length 217;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 LDLLTAV 156
 |||||
 DB 26 LDLLTAV 32

RESULT 24
 B42337
 N:Alternate names: Psp
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Aug-1999
 C:Accession: B42337
 R:Mirels, L.; Ball, W.D.
 J. Biol. Chem. 267, 2679-2687, 1992
 A:Title: Neonatal rat submandibular gland protein SMG-A and parotid secretory protein and parotid secretory protein precursor - rat
 A:Reference number: A42337; MUID:92129360
 A:Accession: B42337
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-235 <MR>
 A:Cross-references: GS:M83209; NID:g206456; PIDN:AAC06334.1; PID:g206457
 A:Note: sequence extracted from NCBI backbone (NCBIN:78709, NCBI:P:78710)
 C:Genetics:
 A:Gene: Psp
 C:Superfamily: parotid secretory protein
 C:Keywords: parotid gland; saliva

Query Match 2.8% Score 7; DB 2; Length 235;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GTSESLL 22
 |||||
 DB 16 GTSESLL 22

RESULT 25
 H97216
 Probable S-adenosylmethionine-dependent methyltransferase [Imported] - *Clostridium acetobutylicum*
 C:Species: *Clostridium acetobutylicum*
 C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: H97216
 R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4623-4638, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: H97216
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-257 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80523.1; PID:g15025597; GSPDB:GNO0168
 A:Experimental source: *Clostridium acetobutylicum* ATCC824
 C:Genetics:
 A:Gene: CAC2574

Query Match 2.8% Score 7; DB 2; Length 257;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 EAEKLLN 86
 |||||
 DB 158 EAEKLLN 164

RESULT 26
 A81504
 conserved hypothetical protein CP1089 [Imported] - *Chlamydomonas reinhardtii* (strain CC-124)
 C:Species: *Chlamydomonas reinhardtii* (strain CC-124)
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: A81504
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hic, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salz, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome Sequences of *Chlamydomonas reinhardtii* and *Chlamydomonas moenchii*
 A:Reference number: A81500; MUID:20150255
 A:Accession: A81504
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-262 <REA>
 A:Cross-references: GB:AE002265; GB:AE002161; NID:g7189999; PIDN:AAF38860.1; PID:g7189999
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP1089

Query Match 2.8% Score 7; DB 2; Length 262;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 KTQLOTL 248
 |||||
 DB 132 KTQLOTL 138

RESULT 27
 B72037
 ct598 hypothetical protein - *Chlamydomonas reinhardtii* (strain CML029)
 C:Species: *Chlamydomonas reinhardtii* (strain CML029)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C:Accession: B72037
 R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of *Chlamydomonas reinhardtii* and *C. trachomatis*.
 A:Reference number: A72000; MUID:99206506
 A:Accession: B72037
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-262 <ARN>
 A:Cross-references: GB:AE001659; GB:AE001363; NID:g4377081; PIDN:AAI18921.1; PID:g4377081
 A:Experimental source: strain CML029
 C:Genetics:
 A:Gene: CP0783

Query Match 2.8% Score 7; DB 2; Length 262;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 KTQLOTL 248
 |||||
 DB 132 KTQLOTL 138

DnaJ-like protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain
 AB05114
 C-species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AB05114
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 et al.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 et al.; Moulie, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AB05114

```

Query Match          2.8%; Score 7; DB 2; Length 271;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 AKQKAQE 80
DB 250 AKQKAQE 256

RESULT 33
D85487
A:Title: probable DNA binding protein yabH [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: D85487
R:Perna N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
illet, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nacure 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85487
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <STO>
A:Cross-references: GB:AF005174; NID:gl2512750; PIDN:AAG54360.1; GSPDB:GN00145; UWGP:200
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yabH
C:Superfamily: dnaJ amino-terminal homology

Query Match          2.8%; Score 7; DB 2; Length 271;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 AKQKAQE 80
DB 250 AKQKAQE 256

RESULT 34
AF0061
A:Title: DnaJ-like protein [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AF0061
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-rarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89353.1; PID:gl15978590; GSPDB:GN00175
C:Genetics:
A:Gene: djIA
C:Superfamily: dnaJ amino-terminal homology

Query Match          2.8%; Score 7; DB 2; Length 277;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 AKQKAQE 80
DB 256 AKQKAQE 262

RESULT 35

```

```

T27690
hypothetical protein ZK1086.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T27690
R:Matthews, L.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20406
A:Accession: T27690
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-290 <WIL>
A:Cross-references: EMBL:Z70269; PIDN:CAA94222.1; GSPDB:GN00028; CESP:ZK1086.3
A:Experimental source: clone ZK1086
C:Genetics:
A:Gene: CESP:ZK1086.3
A:Map position: X
A:Introns: 85/3; 110/1; 125/3; 165/3

Query Match          2.8%; Score 7; DB 2; Length 290;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 ILEKLV 60
DB 142 ILEKLV 148

RESULT 36
T03558
A:Title: ribose transport ATP-binding protein - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Nov-1999
C:Accession: T03558
R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB
A:Reference number: Z14955; MUID:9740404
A:Accession: T03558
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-305 <VLG>
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16211.1; PID:g3128359
C:Genetics:
A:Map position: 1
A:Superfamily: inner membrane protein malK; ATP-binding cassette homology
C:Keywords: ribose transport

Query Match          2.8%; Score 7; DB 2; Length 305;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 LILDVKA 114
DB 193 LILDVKA 199

RESULT 37
S59153
A:Title: NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - land snail mitochondrion
C:Species: mitochondrion Albinaria coerulea (land snail)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Oct-2000
C:Accession: S59153
R:Hatzoglou, E.; Rodakis, G.C.; Lecanidou, R.
Genetics 140, 1353-1366, 1995
A:Title: Complete sequence and gene organization of the mitochondrial genome of the J
A:Reference number: S59143; MUID:96120351
A:Accession: S59153
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-307 <HAT>

```


A:Cross-references: EMBL:X83390; NID:g975668; PIDN:CAA58306.1; PID:g975679; GSPDB:GN0013
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGC4
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2
 C:Keywords: membrane-associated complex; mitochondrion; NAD: oxidative phosphorylation;

Query Match 2.8%; Score 7; DB 2; Length 307;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 SISLSLL 186
 |||||
 Db 218 SISLSLL 224

RESULT 38

AE2384

hypothetical protein alr4629 [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AE2384
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquich
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A:Title: Complete genomic sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AE2384
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-315 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA076328.1; PID:gl7133766; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr4629

Query Match 2.8%; Score 7; DB 2; Length 315;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 SISLSLL 186
 |||||
 Db 6 SISLSLL 12

RESULT 39

G70037

conserved hypothetical protein yvff - Bacillus subtilis

C:Species: Bacillus subtilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: G70037
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
 lech, J.; Harwood, C.R.; Hensut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maues
 Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serof
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tognoni, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: G70037
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-322 <KUN>
 A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PIDN:CAB15427.1; PID:g26359
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yvff
 C:Superfamily: hypothetical protein yxab

Query Match 2.8%; Score 7; DB 1; Length 322;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 SSLLQKE 215
 |||||
 Db 275 SSLLQKE 281

RESULT 40

AH1383

conserved hypothetical protein lmo2472 [imported] - Listeria monocytogenes (strain EG

C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: AH1383
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.;
 ck, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AH1383
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-323 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAD00550.1; PID:gl16411960; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo2472
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCC54.10c

Query Match 2.8%; Score 7; DB 2; Length 323;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GLKISNS 107
 |||||
 Db 317 GLKISNS 323

RESULT 41

AC1920

ABC phosphate transport system permease protein all0910 [imported] - Anabaena sp. (st

C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: AC1920
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriqu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 8, 205-213, 2001
 A:Title: Complete genomic sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AC1920
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-343 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA072867.1; PID:gl7130256; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all0910
 C:Superfamily: phoW protein

Query Match 2.8%; Score 7; DB 2; Length 343;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 ISLSLLD 187
 Db 282 ISLSLLD 288
 |||||

RESULT 42

T13817
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - Atlantic hagfish mitochondrion
 C:Species: mitochondrion Myxine glutinosa (Atlantic hagfish)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
 C:Accession: T13817
 R:Delarbre, C.; Barriel, V.; Tillier, S.; Janvier, P.; Gachelin, G.
 Mol. Biol. Evol. 14, 807-813, 1997
 A:Title: The main features of the craniate mitochondrial DNA between the ND1 and the COI
 A:Reference number: Z17775; MUID:97398704
 A:Accession: T13817
 A:Status: preliminary;
 A:Molecule type: DNA
 A:Residues: 1-348
 A:Cross-references: EMBL:X09527; NID:92340019; PIDN:CAA70717.1; PID:92340021
 C:Genetics:
 A:Genome: mitochondrion
 A:Note: nad2
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2
 C:Keywords: membrane-associated complex, mitochondrion; NAD; oxidative phosphorylation;

Query Match 2.8%; Score 7; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 ISNSLIL 110
 Db 269 ISNSLIL 275
 |||||

RESULT 43

F95861
 hypothetical ABC transporter periplasmic solute-binding protein SMB20158 [imported] - Sinorhizobium meliloti
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: F95861
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: F95861
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-374 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CA48558.1; PID:q15140030; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
 pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMB20158
 A:Genome: Plasmid

Query Match 2.8%; Score 7; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 92;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 93 LPTNTDI 99
 Db 323 LPTNTDI 329
 |||||

RESULT 44

SL8543
 hypothetical protein 5 - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Oct-1999
 C:Accession: SL8543
 R:Caballero, J.L.; Martinez, E.; Malpartida, F.; Hopwood, D.A.
 Mol. Gen. Genet. 230, 401-412, 1991
 A:Title: Organisation and functions of the actA region of the actinorhodin biosynth
 A:Reference number: SL8539; MUID:92114870
 A:Accession: SL8543
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-381 <CAB>
 A:Cross-references: EMBL:X58833; NID:946812; PIDN:CAA41641.1; PID:946817

Query Match 2.8%; Score 7; DB 2; Length 381;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 PVAVLGE 173
 Db 72 PVAVLGE 78
 |||||

RESULT 45

F72352
 hypothetical protein TM0625 - Thermotoga maritima (Strain MSB8)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: F72352
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hi-
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A:Reference number: A72200; MUID:99287316
 A:Accession: F72352
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-383 <ARN>
 A:Cross-references: GB:AE001737; GB:AE000512; NID:94981144; PIDN:AAD35709.1; PID:949
 C:Genetics:
 A:Gene: TM0625
 C:Superfamily: Thermotoga maritima hypothetical protein TM0625

Query Match 2.8%; Score 7; DB 2; Length 383;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 SLDLLTA 155
 Db 184 SLDLLTA 190
 |||||

RESULT 46

DB4068
 RNA helicase BH3348 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: DB4068
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; I
 Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A87650; MUID:20512582; PMID:11058132
 A:Accession: D84068
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-389 <STO>
 A:CROSS-references: GB:AO001518; GB:BA000004; NID:g10175792; PIDN:BA07067.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: B43348
 C:Superfamily: translation initiation factor eIF-4A

Query Match 2.8%; Score 7; DB 2; Length 389;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 KGLNLSF 127
 |||||
 Db 358 KGLNLSF 364

RESULT 47
 B55889
 isoaspartyl dipeptidase (EC 3.4.19.-) - *Escherichia coli*
 C:Species: *Escherichia coli*
 C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 08-Oct-1999
 C:Accession: B55889; S56553; B65247
 R:Gary, J.D.; Clarke, S.
 J. Biol. Chem. 270, 4076-4087, 1995
 A:Title: Purification and characterization of an isoaspartyl dipeptidase from *Escherichia*
 A:Reference number: A55889; MUID:95181377
 A:Accession: B55889
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-390 <GAR>
 A:CROSS-references: GB:U15029; NID:g640029; PIDN:AAC43299.1; PID:g640031
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 23, 2105-2119, 1995
 A:Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.
 A:Reference number: S56314; MUID:95334362
 A:Accession: S56553
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-390 <BUR>
 A:CROSS-references: EMBL:U14003; NID:g1263172; PIDN:AAA97224.1; PID:g537169
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: B65247
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-390 <BLAT>
 A:CROSS-references: GB:AE000503; GB:U00096; NID:g1790777; PIDN:AAC77284.1; PID:g1790784;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: *lady*
 C:Superfamily: *Bacillus* dihydroorotase; *Bacillus* dihydroorotase homology
 C:Keywords: hydrolase; omega peptidase

Query Match 2.8%; Score 7; DB 2; Length 390;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ISKLLPT 95
 |||||
 Db 223 ISKLLPT 229

RESULT 48
 F91289
 isoaspartyl dipeptidase [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001
 C:Accession: F91289
 R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and 9
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: F91289
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-390 <HAY>
 A:CROSS-references: GB:BA000007; PIDN:BA038709.1; PID:g13364764; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: E25286
 C:Superfamily: *Bacillus* dihydroorotase; *Bacillus* dihydroorotase homology

Query Match 2.8%; Score 7; DB 2; Length 390;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ISKLLPT 95
 |||||
 Db 223 ISKLLPT 229

RESULT 49
 C86131
 isoaspartyl dipeptidase [imported] - *Escherichia coli* (strain O157:H7, substrain EDL9
 C:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: C86131
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: C86131
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-390 <STO>
 A:CROSS-references: GB:AE005174; NID:g12519339; PIDN:AAC59511.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: *lady*
 C:Superfamily: *Bacillus* dihydroorotase; *Bacillus* dihydroorotase homology

Query Match 2.8%; Score 7; DB 2; Length 390;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ISKLLPT 95
 |||||
 Db 223 ISKLLPT 229

RESULT 50
 C64553
 chlorohydrolase - *Helicobacter pylori* (strain 26695)
 C:Species: *Helicobacter pylori*
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 17-Mar-2000
 C:Accession: C64553
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
 .Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
 Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,

A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A:Reference number: A64520; MUID:97394467
A:Accession: G64553
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-409 <TOM>
A:Cross-references: GB:AE000545; GB:AE000511; MID:g2313349; PIDN:AAD07333.1; PID:g231335
C:Supertfamily: Aquifex aeolicus N-ethylameline chorohydrolase

Query Match 2.8%; Score 7; DB 2; Length 409;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 181 ISLSLLD 187
|||||
Db 320 ISLSLLD 326

Search completed: August 6, 2002, 16:47:12
Job time: 264 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:46:13 : Search time 10.91 seconds

(without alignments)

883.700 Million cell updates/sec

Title: US-10-020-139-2

Perfect score: 249

Sequence: 1 MQLKWLKVLGCGVLTGTSSES.....NVIQVVDNPNQKTLQTLI 249

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	3.2	561	1 PGMU_HUMAN	P36871 homo sapien
2	8	3.2	561	1 PGMU_RABIT	P00949 oryctolagus
3	8	3.2	561	1 PGMU_RAT	P38652 rattus norv
4	7	2.8	160	1 PHAC_ANASP	P80456 anabaena sp
5	7	2.8	162	1 ATPF_BACCA	P41014 bacillus ce
6	7	2.8	163	1 ATPF_BACP3	P09221 bacillus ps
7	7	2.8	179	1 YCXE_CYPAP	P48335 cyanophora
8	7	2.8	187	1 YQGE_ECOLI	P52049 escherichia
9	7	2.8	207	1 RID2_BUCAI	P57534 buchnera ap
10	7	2.8	216	1 HAN2_CHICK	O90690 gallus gall
11	7	2.8	217	1 HAN2_HUMAN	O95300 homo sapien
12	7	2.8	217	1 HAN2_MOUSE	O61039 m heart- an
13	7	2.8	271	1 DJLA_ECOLI	P31580 escherichia
14	7	2.8	283	1 DAAB_BACLI	P54692 bacillus li
15	7	2.8	307	1 NU2M_ALSCO	P48302 albinaria c
16	7	2.8	348	1 NU2M_MYXGL	O21078 myxine glut
17	7	2.8	390	1 IADA_ECOLI	P39377 escherichia
18	7	2.8	413	1 YV4Q_CAEEL	O45435 caenorhabdi
19	7	2.8	424	1 GCH2_CHLMO	O9pl15 chlamydia m
20	7	2.8	482	1 YQ53_BACAN	O9rmx9 bacillus an
21	7	2.8	588	1 FT51_ECOLI	P04286 escherichia
22	7	2.8	664	1 PLB1_YEAST	P39105 saccharomyc
23	7	2.8	688	1 YJ80_YEAST	P47147 saccharomyc
24	7	2.8	974	1 SIN_YEAST	P32259 saccharomyc
25	7	2.8	1229	1 SIP3_YEAST	P38717 saccharomyc
26	7	2.8	1253	1 SC25_YEAST	P14771 saccharomyc
27	7	2.8	1502	1 N170_YEAST	P38181 saccharomyc
28	7	2.8	1886	1 GP21_RAT	P11654 rattus norv
29	7	2.8	2291	1 SPGB_DROME	Q00963 drosophila
30	7	2.8	2326	1 FGG2_RAT	Q00657 rattus norv
31	6	2.4	12	1 TM2A_METMA	P80652 methanosarc
32	6	2.4	35	1 CECB_ANTPE	P01509 antheraea p
33	6	2.4	68	1 TXC9_CUPSA	P58604 cupliennius

34	6	2.4	78	1 RP5M_THIFE	P24694 thiobacillu
35	6	2.4	80	1 SSS2_SCYCA	P11020 scylliorhinu
36	6	2.4	81	1 Y039_SYNP7	P27368 synechococc
37	6	2.4	87	1 Y020_MYCPN	P75237 mycoplasma
38	6	2.4	90	1 ICBR_HUMAN	P57730 homo sapien
39	6	2.4	91	1 YP98_CAEEL	O29946 caenorhabdi
40	6	2.4	92	1 Y249_ARCFU	O29990 archaeoglob
41	6	2.4	95	1 MYLE_HUMAN	O95424 homo sapien
42	6	2.4	95	1 MYLE_MOUSE	O95407 mus musculu
43	6	2.4	96	1 FSPM_LYCES	P14903 lycopersico
44	6	2.4	97	1 LTUB_CHLTR	Q46404 chlamydia t
45	6	2.4	97	1 YAFN_ECOLI	Q47156 escherichia
46	6	2.4	98	1 C552_HYDTH	P15452 hydrogenoba
47	6	2.4	101	1 RK24_GUITH	O46905 guillardia
48	6	2.4	104	1 RFA3_SCHPO	Q92374 schizosacch
49	6	2.4	109	1 Y20L_SYNY3	P72983 synechocyst
50	6	2.4	116	1 HV01_HETFR	P03983 heterodontu
51	6	2.4	119	1 V717_METJA	Q58127 methanococc
52	6	2.4	122	1 RK14_ASTJO	P58139 astasia ion
53	6	2.4	123	1 UR2_RAT	Q9qz94 rattus norv
54	6	2.4	125	1 MFAP_SACKL	P06648 saccharomyc
55	6	2.4	129	1 VDBP_CAMVC	P03552 cauliflower
56	6	2.4	129	1 VDBP_CAMVD	P03553 cauliflower
57	6	2.4	129	1 VDBP_CAMVE	Q02967 cauliflower
58	6	2.4	129	1 VDBP_CAMVN	Q00967 cauliflower
59	6	2.4	129	1 VDBP_CAMVS	P03551 cauliflower
60	6	2.4	130	1 RK12_PROWI	O99010 prototheca
61	6	2.4	132	1 TVA3_MOUSE	P06323 mus musculu
62	6	2.4	141	1 REG7_PYRFU	P42180 pyrococcus
63	6	2.4	141	1 REG7_PYRHO	O59256 pyrococcus
64	6	2.4	142	1 NIJ2_HUMAN	O9n297 homo sapien
65	6	2.4	145	1 DASA_HAEIN	P43758 haemophilus
66	6	2.4	146	1 CYNS_SYNP7	Q59948 synechococc
67	6	2.4	149	1 CYNS_SYNY3	Q55367 synechocyst
68	6	2.4	149	1 TRJ8_ECOLI	P33786 escherichia
69	6	2.4	150	1 GAG_AVISU	P03324 avian sarco
70	6	2.4	150	1 MLE_DICDI	P09402 dictyosteli
71	6	2.4	150	1 RL9_MYCPN	Q98q52 mycoplasma
72	6	2.4	157	1 Y254_MYCPN	Q9exq9 mycoplasma
73	6	2.4	162	1 MCRW_METTH	Q50485 methanobact
74	6	2.4	167	1 F6B1_ECOLI	P53510 escherichia
75	6	2.4	167	1 F6B2_ECOLI	P53511 escherichia

ALIGNMENTS

RESULT 1
PGMU_HUMAN
ID PGMU_HUMAN STANDARD; PRT: 561 AA.
AC P36871: Q16105; Q9NTY4;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoglucosyltransferase (EC 5.4.2.2) (Glucose phosphomutase) (PGM).

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=92108065; PubMed=1530890;
RA Whitehouse D.B., Putt W., Lovegrove J.U., Morrison K.E., Hollyoake M.,
RA Fox M.F., Hopkinson D.A., Edwards Y.H.;
RT "Phosphoglucosyltransferase 1: complete human and rabbit mRNA sequences and
RT direct mapping of this highly polymorphic marker on human chromosome
RT 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:411-415(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Donnelly S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RT "The crystal structure of muscle phosphoglucumutase refined at 2.7-A
 resolution.";
 RL J. Biol. Chem. 267:6322-6337(1992).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RA Liu Y., Ray W.J. Jr., Baranidharan S.:
 RT "Structure of rabbit muscle phosphoglucumutase refined at 2.4-A
 resolution.";
 RL Acta Crystallogr. D 53:392-405(1997).
 CC -!- FUNCTION: THIS ENZYME PARTICIPATES IN BOTH THE BREAKDOWN AND
 CC SYNTHESIS OF GLUCOSE.
 CC -!- CATALYTIC ACTIVITY: Alpha-D-glucose 1-phosphate -> alpha-D-glucose
 CC 6-phosphate.
 CC -!- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
 CC -!- SUBUNIT: MONOMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (ISOFORM 1); JUNCTIONAL SKELETAL
 CC SARCOPLASMIC RETICULUM, PROBABLY BY ASSOCIATION WITH PHOSPHOLIPIDS
 CC AND/OR OTHER PROTEINS (ISOFORM 2).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- PTM: ISOFORM 2 IS THE MAJOR CALMODULIN-DEPENDENT PHOSPHOPROTEIN IN
 CC JUNCTIONAL SKELETAL SARCOPLASMIC RETICULUM VESICLES.
 CC -!- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.
 CC -!- DATABASE: NAME=Worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/P/PGM.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M97664; AAA31454.1; .
 DR EMBL: M97663; AAA31453.1; .
 DR PIR: A01175; PMRB.
 DR PIR: A45077; PMRB1.
 DR PIR: B41801; B41801.
 DR PDB: 3PMG; 07-DEC-95.
 DR PDB: 1UDY; 12-MAR-97.
 DR PDB: 1LXI; 12-FEB-97.
 DR PDB: 1VKL; 11-JAN-97.
 DR InterPro: IPR001485; PGM_PMM.
 DR Pfam: PF00408; PGM_PMM; 1.
 DR Pfam: PF02878; PGM_PMM.I; 1.
 DR Pfam: PF02879; PGM_PMM.II; 1.
 DR Pfam: PF02880; PGM_PMM.III; 1.
 DR PRINTS: PR00509; PGM_PMM.
 DR PROSITE: PS00710; PGM_PMM; 1.
 KW Isomerase; Phosphorylation; Magnesium; Alternative splicing;
 KW 3D-structure.
 FT INIT_MET 0 0
 FT ACT_SITE 116 116 FORMS THE PHOSPHOSERINE INTERMEDIATE.
 FT VARSPIC 1 12 VKIVVTKAYP->EEGLPLLTITATPYH (IN
 FT ISOFORM 2).
 FT VARSPIC 24 35 RKVVFQSTNVA->KTYVFEDKPCYL (IN ISOFORM
 FT 2).
 FT VARSPIC 43 55 ISTVEPAQRQKAT->FFSIDLKDRGSS (IN
 FT ISOFORM 2).
 FT VARSPIC 64 76 FYMKEAQLIVRI->YFNKSALETILQM (IN
 FT ISOFORM 2).
 FT STRAND 4 6
 FT TURN 17 17
 FT TURN 18 19
 FT STRAND 20 24
 FT TURN 25 30
 FT TURN 32 33
 FT TURN 34 44
 FT TURN 45 45
 FT TURN 48 49
 FT TURN 50 53
 FT STRAND 55 60

FT TURN 64 65
 FT TURN 66 80
 FT TURN 81 81
 FT STRAND 84 92
 FT TURN 95 105
 FT TURN 106 106
 FT STRAND 109 113
 FT TURN 116 117
 FT TURN 122 123
 FT STRAND 125 132
 FT TURN 133 134
 FT STRAND 137 137
 FT TURN 140 150
 FT TURN 151 152
 FT STRAND 156 158
 FT TURN 166 167
 FT STRAND 170 175
 FT TURN 177 178
 FT STRAND 182 188
 FT TURN 192 199
 FT TURN 200 201
 FT TURN 204 212
 FT TURN 214 215
 FT STRAND 221 222
 FT TURN 224 225
 FT TURN 227 228
 FT TURN 229 239
 FT TURN 240 240
 FT TURN 244 246
 FT STRAND 247 248
 FT TURN 267 268
 FT TURN 269 275
 FT TURN 276 277
 FT STRAND 283 286
 FT TURN 293 296
 FT TURN 298 300
 FT STRAND 302 303
 FT TURN 305 313
 FT TURN 314 315
 FT TURN 316 318
 FT TURN 320 325
 FT STRAND 330 333
 FT TURN 334 335
 FT TURN 338 344
 FT TURN 345 346
 FT STRAND 351 354
 FT TURN 358 365
 FT TURN 366 368
 FT TURN 372 375
 FT TURN 376 378
 FT TURN 379 382
 FT TURN 383 384
 FT TURN 390 403
 FT TURN 404 405
 FT TURN 408 419
 FT STRAND 421 429
 FT STRAND 435 440
 FT STRAND 448 456
 FT TURN 460 463
 FT TURN 464 464
 FT TURN 465 473
 FT TURN 474 475
 FT STRAND 486 490
 FT TURN 496 501
 FT TURN 515 516
 FT TURN 517 524
 FT STRAND 532 537
 FT STRAND 540 541
 FT STRAND 548 549
 SQ SEQUENCE 561 AA: 61427 MW: 608AFB26B0242832 CRC64;

Query Match

3.28; Score 8; DB 1; Length 561;

Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 LKVDLGLV 65
Db 162 LKVDLGLV 169

RESULT 3
PGMU_RAT
ID PGMU_RAT STANDARD; PRT; 561 AA.
AC P38652;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoglucosyltransferase (EC 5.4.2.2) (Glucose phosphomutase) (PGM).
GN PGM1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94040821; PubMed=8224913;
RA Rivera A.A., Eilon T.S., Day N.B., Bounelis P., Marchase R.B.;
RT "Isolation and expression of a rat liver cDNA encoding
RT phosphoglucosyltransferase";
RL Gene 133:261-266(1993).
CC -1- FUNCTION: THIS ENZYME PARTICIPATES IN BOTH THE BREAKDOWN AND
CC SYNTHESIS OF GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Alpha-D-glucose 1-phosphate = alpha-D-glucose
CC 6-phosphate.
CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L11694; AAA16862.1; -
DR PIR: JC2011; PMWT.
DR HSP: P00949; 3PMG.
DR InterPro: IPR001485; PGM_PMM.
DR Pfam: PF00408; PGM_PMM; 1.
DR Pfam: PF02878; PGM_PMM_1; 1.
DR Pfam: PF02879; PGM_PMM_11; 1.
DR Pfam: PF02880; PGM_PMM_111; 1.
DR PRINTS: PR00509; PGM_PMM.
DR PROSITE: PS00710; PGM_PMM; 1.
KW Isomerase; Phosphorylation; Magnesium.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 116 116 FORMS THE PHOSPHOSERINE INTERMEDIATE
FT (BY SIMILARITY).
SQ SEQUENCE 561 AA; 61272 MW; 585A151D49170A66 CRC64;

Query Match 3.2%; Score 8; DB 1; Length 561;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 LKVDLGLV 65
Db 162 LKVDLGLV 169

RESULT 4

PHAC_ANASP
ID PHAC_ANASP STANDARD; PRT; 160 AA.
AC P80556;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Allophycocyanin alpha-B chain.
GN APCD OK ALL3653.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120";
RL DNA Res. 8:205-213(2001).
RN [2]
RP SEQUENCE OF 1-28.
RX MEDLINE=96270757; PubMed=8655889;
RA Ducret A., Sidler W., Wehrli E., Frank G., Zuber H.;
RT "Isolation, characterization and electron microscopy analysis of a
RT hemidiscoidal phycobilisome type from the cyanobacterium Anabaena sp.
RL PCC 7120";
RL Eur. J. Biochem. 236:1010-1024(1996).
CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
CC FROM THE PHYCOBILIPROTEIN COMPLEX. ALLOPHYCOCYANIN HAS A MAXIMUM
CC ABSORPTION AT APPROXIMATELY 654 NANOMETERS.
CC -1- SUBUNIT: HETEROHETEROMER OF TWO ALPHA CHAINS, ONE ALPHA-B CHAIN AND
CC THREE BETA CHAINS.
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP003593; BAB7352.1; -
DR InterPro: IPR001659; Phycobilisome.
DR Pfam: PF00502; Phycobilisome; 1.
KW Phycobilisome; Electron transport; Photosynthesis; Bile pigment;
KW Methylation; Complete proteome.
FT INIT_MET 0 0
FT MOD_RES 70 70 METHYLATION (BY SIMILARITY).
FT BINDING 80 80 PHYCOCYANOBILIN CHROMOPHORE (BY
FT SIMILARITY).
SQ SEQUENCE 160 AA; 17680 MW; 5E814C32F2857BD6 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 160;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 KASLDLL 153
Db 134 KASLDLL 140

RESULT 5
ATPF_BACCA
ID ATPF_BACCA STANDARD; PRT; 162 AA.
AC P41014;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)


```

DE  ATP synthase B chain (EC 3.6.3.14).
GN  ATPF
OS  Bacillus caldotenax.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group;
OC  Bacillus/Staphylococcus group; Bacillus.
OX  NCBI_TaxID=1395;
RN  [1]
RA  SEQUENCE FROM N.A.
RL  Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC  -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC  CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC  SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC  HAS THREE MAIN SUBUNITS: A, B AND C.
CC  -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: D38057; BAA07245.1;
DR  InterPro: IPR002146; ATP-synt_B.
DR  Pfam: PF00430; ATP-synt_B; 1.
KW  Hydrogen ion transport; Transmembrane; CF(0).
FT  TRANSMEM 16 36 POTENTIAL.
SQ  SEQUENCE 162 AA; 18598 MW; AC75BF93D031C088 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 QEAEKLL 85
DB 57 QEAEKLL 73

RESULT 6
ATPF_BACP3 STANDARD; PRT: 163 AA.
AC P09221;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP synthase B chain precursor (EC 3.6.3.14).
GN ATPF.
OS Bacillus PS3 (Thermophilic bacterium PS-3).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=70306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=68163679; PubMed=2894854;
RA Ohta S., Yoshida M., Ishizuka M., Hirata H., Hamamoto T.,
RA Otawara-Hamamoto Y., Matsuda K., Kagawa Y.;
RT "Sequence and over-expression of subunits of adenosine triphosphate
RT synthase in thermophilic bacterium PS3.;"
RL Biochim. Biophys. Acta 933:141-155(1988).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DE  ATP synthase B chain (EC 3.6.3.14).
GN  ATPF
OS  Bacillus caldotenax.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group;
OC  Bacillus/Staphylococcus group; Bacillus.
OX  NCBI_TaxID=1395;
RN  [1]
RA  SEQUENCE FROM N.A.
RL  Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC  -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC  CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC  SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC  HAS THREE MAIN SUBUNITS: A, B AND C.
CC  -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: X07804; CAA30650.1;
DR  PIR: S01399; S01399.
DR  InterPro: IPR002146; ATP-synt_B.
DR  Pfam: PF00430; ATP-synt_B; 1.
KW  Hydrogen ion transport; Transmembrane; CF(0).
FT  PROSP 1 11
FT CHAIN 12 163 ATP SYNTHASE B CHAIN.
FT TRANSMEM 16 36 POTENTIAL.
SQ  SEQUENCE 163 AA; 18566 MW; B8E790CA0EF18ACF CRC64;

Query Match 2.8%; Score 7; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 QEAEKLL 85
DB 67 QEAEKLL 73

RESULT 7
YCXE_CYAPA STANDARD; PRT: 179 AA.
ID YCXE_CYAPA
AC P48335;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 21.2 kDa protein in YCF23-APCF intergenic region
DE (ORF179).
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / PRINGSHEIM;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.;"
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / PRINGSHEIM;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.;"
RL (In) Schenk H.F.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwenmiller W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: U30821; AAA81305.1;
DR  Hypothetical protein; Cyanelle.
SQ  SEQUENCE 179 AA; 21248 MW; 046AF1943F4F57E2 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 165 HOPVAVL 171
DB 7 HOPVAVL 13

RESULT 8
YQGE_ECOLI
ID YQGE_ECOLI STANDARD: PRT: 187 AA.
AC P52049; P76648;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yqge.
GN YQGE OR B2948.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)_TaxID=562;
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN (2)
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=99420866; PubMed=10493123;
RA Fountoulakis M., Takacs M.F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxamate chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE HI0304.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U28377; AAA69115.1; ALT_INIT.
DR EMBL; AE000377; AAC75985.1; ALT_INIT.
DR EcoGene; EG13289; YQGE
DR InterPro; IPR003774; DUF179.
DR Pfam; PF02622; DUF179; 1.
KW Complete proteome.
SQ SEQUENCE 187 AA: 20686 MW: 84000CACAPHF4E025 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 187;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 53 GILEKLK 59
DB 50 GILEKLK 56

RESULT 9
RID2_BUCAI
ID RID2_BUCAI STANDARD: PRT: 207 AA.
AC P57534;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 5-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193) (HTP
DE reductase).
GN RIBD2 OR BU0462.

```

```

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118699;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: 5-amino-6-(5-phosphoribitylamino)uracil +
CC NADP(+) = 5-amino-6-(5-phosphoribosylamino)uracil + NADPH.
CC -!- PATHWAY: THIRD STEP OF RIBOFLAVIN BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE HTP REDUCTASE FAMILY.
CC -!- CAUTION: It is possible that ribd1 and ribd2 are, like in all
CC other bacterial species, encoded on one gene and that there is a
CC sequencing error that has produced 2 ORfs instead of one.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001119; BAB13159.1;
DR InterPro; IPR002734; RIBD_C.
DR Pfam; PF01872; RIBD_C; 1.
KW Riboflavin biosynthesis; Oxidoreductase; NADP: Complete proteome.
SQ SEQUENCE 207 AA: 23942 MW: 240973284089576A CRC64;

Query Match 2.8%; Score 7; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 102 LKTSNLSL 108
DB 178 LKTSNLSL 184

RESULT 10
HAN2_CHICK
ID HAN2_CHICK STANDARD: PRT: 216 AA.
AC Q90690;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heart- and neural crest derivatives-expressed protein 2 (Deciduum,
DE heart, autonomic nervous system and neural crest derivatives-
DE expressed protein 2) (dHAND).
GN HAN2 OR dHAND.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)_TaxID=9031;
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=9613273; PubMed=8533092;
RA Srivastava D., Cserjesi P., Olson E.N.;
RT "A subclass of bHLH proteins required for cardiac morphogenesis.";
RL Science 270:1995-1998(1995).
RN (2)
RP FUNCTION, AND DEVELOPMENTAL STAGE.
RX MEDLINE=20233834; PubMed=10769237;
RA Fernandez-Teran M., Piedra M.E., Kathirya I.S., Srivastava D.,
RA Rodriguez-Rey J.C., Ros M.A.;
RT "Role of dHAND in the anterior-posterior polarization of the limb bud;

```

RT implications for the sonic hedgehog pathway.";

RL Development 127:2133-2142(2000).

CC -!- FUNCTION: ESSENTIAL FOR CARDIAC MORPHOGENESIS. BINDS DNA ON E-BOX

CC CONSENSUS SEQUENCE 5'-CANNTG-3' (BY SIMILARITY). PLAYS AN

CC IMPORTANT ROLE IN LIMB DEVELOPMENT, PARTICULARLY IN THE

CC ESTABLISHMENT OF ANTERIOR-POSTERIOR POLARIZATION OF THE LIMB BUD.

CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER

CC BHLH PROTEIN.

CC -!- DEVELOPMENTAL STAGE: AT STAGES 8 TO 9, EXPRESSED IN THE WHOLE

CC LATERAL PLATE MESODERM AND PRECARDIOGENIC MESODERM. AT STAGE 10,

CC EXPRESSED THROUGHOUT THE CARDIAC TUBE AND THE SINUS VENOSUS. BY

CC STAGE 15, EXPRESSED HOMOGENEUSLY IN THE VARIOUS REGION OF THE

CC HEART, INCLUDING THE ATRIA, FUTURE LEFT VENTRICLE, BULBUS CORDIS

CC AND TRUNCUS ARTERIOSUS, AND IN THE FORMING BRANCHIAL ARCHES.

CC EXPRESSION PERSISTS THROUGH STAGE 20, BUT DECREASES THEREAFTER. IN

CC THE DEVELOPING LIMBS, FROM THE INITIATION OF THE BUDS (STAGES 16

CC TO 17), EXPRESSION IS DOWN-REGULATED AT THE ANTERIOR OF THE LIMB

CC BUDS SO THAT A GRADIENT REGRESSION ALONG THE ANTERIOR-POSTERIOR

CC AXIS OF THE BUD IS ESTABLISHED WITH HIGHER EXPRESSION AT THE

CC POSTERIOR BORDER. AT LATER STAGES, EXPRESSION IS RESTRICTED TO THE

CC POSTERIOR BORDER OF THE ZEUGPOD AND TO THE POSTERIOR AUTOPOD. IN

CC THE AUTOPOD, DYNAMIC EXPRESSION OF HAND2 AFFECTS THE INTERDIGITAL

CC REGIONS, THE LATERAL BORDERS OF THE DIGITS AND EVENTUALLY THE

CC DEVELOPING VENTRAL TENDONS.

CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

CC TRANSCRIPTION FACTORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U40040; AAC59733.1; -

DR HSSP: P10085; IMDY.

DR InterPro: IPR003015; HLH_Myc.

DR InterPro: IPR001092; HLH_dim.

DR Pfam: PF00010; HLH; 1.

DR SMART: SM00353; HLH; 1.

DR PROSITE: PS00038; HELIX_LOOP_HELIX; FALSE_NEG.

DR Developmental protein.

DR Nuclear protein.

CC

FW DOMAIN 21 30 POLY-ARG.

FT DNA_BIND 98 110 BASIC DOMAIN.

FT DOMAIN 111 154 HELIX-LOOP-HELIX MOTIF.

FW SEQUENCE 216 AA; 24407 MW; 0B3F052F2CC34C6B CRC64;

Query Match 2.8%; Score 7; DB 1; Length 216;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 LKSTVSS 210

Db 188 LKSTVSS 194

|||||||

RESULT 11

HAN2_HUMAN STANDARD; PRT; 217 AA.

ID HAN2_HUMAN

AC O95300; O95301; P97833.

DT 16-OCT-2001 (Rel. 40; Created)

DT 16-OCT-2001 (Rel. 40; Last sequence update)

DE 01-MAR-2002 (Rel. 41; Last annotation update)

DE Heart- and neural crest derivatives-expressed protein 2 (Deciduum,

DE heart, autonomic nervous system and neural crest derivatives-

DE expressed protein 2) (dHAND).

GN HAND2 OR dHAND.

OS Homo sapiens (Human), and

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606, 10116;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=Human; TISSUE=Embryonic heart;

RX MEDLINE=99096473; PubMed=9878849;

RA Russell M.W., Kemp P., Wang L., Brody L.C., Izumo S.;

RT "Molecular cloning of the human HAND2 gene.";

RL Biochim. Biophys. Acta 1443:393-399(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=Rat; STRAIN=Wistar; TISSUE=Aorta;

RA Kemp P., Grainger D., Metcalfe J.;

RT "Expression of helix-loop-helix proteins in vascular smooth muscle.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: ESSENTIAL FOR CARDIAC MORPHOGENESIS, PARTICULARLY FOR

CC THE FORMATION OF THE RIGHT VENTRICLE AND OF THE AORTIC ARCH

CC ARTERIES. REQUIRED FOR VASCULAR DEVELOPMENT AND REGULATION OF

CC ANGIOGENESIS, POSSIBLY THROUGH A VEGF SIGNALING PATHWAY. PLAYS

CC ALSO AN IMPORTANT ROLE IN LIMB DEVELOPMENT, PARTICULARLY IN THE

CC ESTABLISHMENT OF ANTERIOR-POSTERIOR POLARIZATION, ACTING AS AN

CC UPSTREAM REGULATOR OF SONIC HEDGEHOG (SHH) INDUCTION IN THE LIMB

CC BUD. IS INVOLVED IN THE DEVELOPMENT OF BRANCHIAL ARCHES, WHICH

CC GIVE RISE TO UNIQUE STRUCTURES IN THE HEAD AND NECK. BINDS DNA ON

CC E-BOX CONSENSUS SEQUENCE 5'-CANNTG-3' (BY SIMILARITY).

CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER

CC BHLH PROTEIN. FORMS HOMODIMERS AND HETERODIMERS WITH TCF3 GENE

CC PRODUCTS E12 AND E47, HAND1 AND HEY1, HEY2 AND HEYL (HAIRY-RELATED

CC TRANSCRIPTION FACTORS) (BY SIMILARITY).

CC -!- TISSUE SPECIFICITY: HEART.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE FETAL HEART.

CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

CC TRANSCRIPTION FACTORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF087940; AAD13185.1; -

DR EMBL: AF087941; AAD13186.1; -

DR EMBL: Y08138; CAA69332.1; -

DR HSSP: P10085; IMDY.

DR MIM: 602407; -

DR InterPro: IPR003015; HLH_Myc.

DR InterPro: IPR001092; HLH_dim.

DR Pfam: PF00010; HLH; 1.

DR SMART: SM00353; HLH; 1.

DR PROSITE: PS00038; HELIX_LOOP_HELIX; FALSE_NEG.

DR Developmental protein; Angiogenesis; Transcription regulation;

FW DNA-binding; Nuclear protein.

FT DOMAIN 21 32 POLY-ALA.

FT DNA_BIND 99 111 BASIC DOMAIN.

FT DOMAIN 112 155 HELIX-LOOP-HELIX MOTIF.

FW SEQUENCE 217 AA; 23666 MW; 528F541BB2173F1E CRC64;

Query Match 2.8%; Score 7; DB 1; Length 217;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 LKSTVSS 210

Db 189 LKSTVSS 195

|||||||

RESULT 12

HAN2_MOUSE STANDARD; PRT; 217 AA.

ID HAN2_MOUSE

AC Q61039; Q61100;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Heart- and neural crest derivatives-expressed protein 2 (Deciduum,
 DE heart, autonomic nervous system and neural crest derivatives-
 DE expressed protein 2) (dHAND) (Helix-loop-helix transcription factor
 DE expressed in embryo and deciduum) (Thing-2).
 GN HAND2 OR dHAND OR HED OR THING2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=96123273; PubMed=8533092;
 RA Srivastava D., Cserjesi P., Olson E.N.;
 RT "A subclass of bHLH proteins required for cardiac morphogenesis.";
 RL Science 270:1995-1998(1995).
 RN [2]
 RP SEQUENCE OF 92-203 FROM N.A.
 RC STRAIN=129/SV; TISSUE=Embryoid bodies;
 RX MEDLINE=95401868; PubMed=7671815;
 RA Cross J.C., Flannery M.L., Blannar M.A., Steingrimsen E.,
 RA Jenkins N.A., Copeland N.G., Rutter W.J., Werb Z.;
 RT "Hxt encodes a basic helix-loop-helix transcription factor that
 RT regulates trophoblast cell development.";
 RL Development 121:2513-2523(1995).
 RN [3]
 RP FUNCTION, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=97315821; PubMed=9171826;
 RA Srivastava D., Thomas T., Lin Q., Kirby M.L., Brown D., Olson E.N.;
 RT "Regulation of cardiac mesodermal and neural crest development by the
 RT bHLH transcription factor, dHAND.";
 RL Nat. Genet. 16:154-160(1997).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=98337821; PubMed=9671575;
 RA Thomas T., Kurihara H., Yamagishi H., Kurihara Y., Yazaki Y.,
 RA Olson E.N., Srivastava D.;
 RT "A signaling cascade involving endothelin-1, dHAND and msxl regulates
 RT development of neural crest-derived branchial arch mesenchyme.";
 RL Development 125:3005-3014(1998).
 RN [5]
 RP SUBUNIT.
 RX MEDLINE=20517917; PubMed=10924525;
 RA Firulli B.A., Hadzic D.B., McDavid J.R., Firulli A.B.;
 RT "The basic helix-loop-helix transcription factors dHAND and eHAND
 RT exhibit dimerization characteristics that suggest complex regulation
 RT of function.";
 RL J. Biol. Chem. 275:33567-33573(2000).
 RN [6]
 RP FUNCTION, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=20265920; PubMed=10804186;
 RA Charite J., McFadden D.G., Olson E.N.;
 RT "The bHLH transcription factor dHAND controls Sonic hedgehog
 RT expression and establishment of the zone of polarizing activity
 RT during limb development.";
 RL Development 127:2461-2470(2000).
 RN [7]
 RP FUNCTION, AND DEVELOPMENTAL STAGE.
 RX MEDLINE=20233834; PubMed=10769237;
 RA Fernandez-Teran M., Piedra M.E., Kathiraya I.S., Srivastava D.,
 RA Rodriguez-Rey J.C., Ros M.A.;
 RT "Role of dHAND in the anterior-posterior polarization of the limb bud:
 RT implications for the Sonic hedgehog pathway.";
 RL Development 127:2133-2142(2000).
 CC -!- FUNCTION: ESSENTIAL FOR CARDIAC MORPHOGENESIS, PARTICULARLY FOR
 CC THE FORMATION OF THE RIGHT VENTRICLE AND OF THE AORTIC ARCH
 CC ARTERIES. REQUIRED FOR VASCULAR DEVELOPMENT AND REGULATION OF
 CC ANGIOGENESIS, POSSIBLY THROUGH A VEGF SIGNALING PATHWAY. PLAYS
 CC ALSO AN IMPORTANT ROLE IN LIMB DEVELOPMENT, PARTICULARLY IN THE
 CC ESTABLISHMENT OF ANTERIOR-POSTERIOR POLARIZATION, ACTING AS AN
 CC UPSTREAM REGULATOR OF SONIC HEDGEHOG (SHH) INDUCTION IN THE LIMB
 CC BUD. IS INVOLVED IN THE DEVELOPMENT OF BRANCHIAL ARCHES, WHICH
 CC GIVE RISE TO UNIQUE STRUCTURES IN THE HEAD AND NECK. BINDS DNA ON
 CC E-BOX CONSENSUS SEQUENCE 5'-CANNIG-3'.
 CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC bHLH PROTEIN. FORMS HOMODIMERS AND HETERODIMERS WITH TCF3 GENE
 CC PRODUCTS E12 AND E47, HAND1 AND HEY1, HEV2 AND HEYL (HAIRY-RELATED
 CC TRANSCRIPTION FACTORS).
 CC -!- TISSUE SPECIFICITY: HEART AND AORTA.
 CC -!- DEVELOPMENTAL STAGE: HIGH EXTREME EMBRYONIC EXPRESSION IS DETECTED AT
 CC 7.5 DPC IN THE MATERNALLY DERIVED DECIDUUM. ALSO DETECTED ALONG
 CC THE YOLK SAC VESSELS DURING THE PROCESS OF REMODELING AT 9.5-10.0
 CC DPC. WITHIN THE EMBRYO, DETECTED AT 7.5 DPC IN THE LATERAL
 CC MESODERM INCLUDING THE PRECARDIAC MESODERM. ON DAY 8.5 PC
 CC EXPRESSED THROUGHOUT THE STRAIGHT HEART TUBE. IN THE CAUDAL REGION
 CC OF THE EMBRYO, EXPRESSED IN THE LATERAL MESODERM AT THE LEVEL OF
 CC SEPARATION OF THE SOMATIC AND Splanchnic Mesoderm. ON DAY 9.5 PC
 CC EXPRESSED THROUGHOUT THE DEVELOPING CARDIOVASCULAR REGION, MOST
 CC ABUNDANT IN THE OUTFLOW TRACT AND IN THE FIRST AND SECOND AORTIC
 CC ARCH ARTERIES, AND IN PHARYNGEAL ARCHES. AS THE HEART LOOPS, THE
 CC EXPRESSION BECOMES RESTRICTED TO THE CONOTRUNCUS AND FUTURE RIGHT
 CC VENTRICLE (ENDOCARDIUM AND MYOCARDIUM). AT 10.5 DPC, HIGHLY
 CC EXPRESSED IN THE BRANCHIAL ARCHES, AS WELL AS IN THE TRUNCUS
 CC ARTERIOSUS, AORTIC SAC, AND THE VASCULAR MESENCHYME BETWEEN THE
 CC THIRD AND FOURTH AORTIC ARCH ARTERIES, WHICH LATER GIVES RISE TO
 CC VASCULAR SMOOTH MUSCLE CELLS AND TO THE MESENCHYME OF THE
 CC PHARYNGEAL ARCH. ON DAY 13.5 PC, BARELY DETECTABLE IN THE HEART,
 CC BUT APPARENT IN THE NEURAL CREST-DERIVED SYMPATHETIC TRUNK AND
 CC ADRENAL MEDULLA, A PATTERN SIMILAR TO THAT OF HAND1. IN THE
 CC DEVELOPING LIMBS, EXPRESSION IS DETECTED IN THE POSTERIOR MESODERM
 CC OF THE BUDS AT 9.5 DPC. IT IS THEN PROGRESSIVELY DOWN-REGULATED AT
 CC THE ANTERIOR-POSTERIOR AXIS OF THE BUD SO THAT A GRADIENT EXPRESSION ALONG
 CC THE ANTERIOR-POSTERIOR BORDER. AT LATER STAGES OF LIMB
 CC DEVELOPMENT, EXPRESSION IS RESTRICTED TO THE POSTERIOR BORDER OF
 CC THE ZEUGOPOD AND TO THE POSTERIOR AUTOPOD. IN THE AUTOPOD, DYNAMIC
 CC EXPRESSION OF HAND2 AFFECTS THE INTERDIGITAL REGIONS, THE LATERAL
 CC BORDERS OF THE DIGITS AND EVENTUALLY THE DEVELOPING VENTRAL
 CC TENDONS. AFTER DAY 16 PC, EXPRESSION DECREASES THROUGHOUT THE
 CC EMBRYO.
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (bHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS.
 CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 183
 CC ONWARD DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U40039; AAC52338.1; -
 CC EMBL; U43715; AAA86274.1; ALT_FRAME.
 CC HSSP; P10085; IMDY.
 CC MGD; MG1:103580; Hand2.
 CC InterPro; IPR003015; HLH_Myc.
 CC InterPro; IPR001092; HLH_dim.
 CC Pfam; PF00010; HLH; 1.
 CC SMART; SM00353; HLH; 1.
 CC PROSITE; PS00038; HELIX_LOOP_HELIX; FALSE_NEG.
 CC Developmental protein; Angiogenesis; Transcription regulation;
 CC DNA-binding; Nuclear protein.
 CC DOMAIN 29 32 POLY-HIS.
 CC DNA_BIND 99 111 BASIC DOMAIN.
 CC DOMAIN 112 155 HELIX-LOOP-HELIX MOTIF.
 CC CONFLICT 92 93 LG -> TA (IN REF. 2).
 CC CONFLICT 156 156 A -> P (IN REF. 1).
 CC CONFLICT 166 166 A -> G (IN REF. 2).
 CC SEQUENCE 217 AA; 24349 MW; 5B4F0AF5DF367115 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 217;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 LKSTVSS 210
 |||||
 DB 189 LKSTVSS 195

RESULT 13
 DJLA_ECOLI STANDARD; PRT: 271 AA.
 AC P31680;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DnaJ-like protein djla.
 GN DJLA OR B0055.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-92334977; PubMed-1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RA Isono K., Mizobuchi K., Nakata A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 0-2.4 min region."
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Kelley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1234-1238(1997).
 RN [3]
 RP CHARACTERIZATION, AND SEQUENCE OF 1-45 FROM N.A.
 RX MEDLINE-96405649; PubMed-8809778;
 RA Clarke D.J., Jacq A., Holland I.B.;
 RT "A novel DnaJ-like protein in Escherichia coli inserts into the
 cytoplasmic membrane with a type III topology."
 RL Mol. Microbiol. 20:1273-1286(1996).
 RN [4]
 RP CHARACTERIZATION, AND MUTAGENESIS OF HIS-233.
 RX MEDLINE-98030199; PubMed-9364917;
 RA Kelley W.L., Georgopoulos C.;
 RT "Positive control of the two-component RnsC/B signal transduction
 network by DnaJ: a member of the DnaJ family of molecular chaperones
 in Escherichia coli."
 RL Mol. Microbiol. 25:933-944(1997).
 RN [5]
 RP MUTAGENESIS.
 RX MEDLINE-98030200; PubMed-9364918;
 RA Clarke D.J., Holland I.B., Jacq A.;
 RT "Point mutations in the transmembrane domain of DnaJ, a
 membrane-linked DnaJ-like protein, abolish its function in promoting
 colanic acid production via the Rcs signal transduction pathway."
 RL Mol. Microbiol. 25:933-944(1997).
 CC -1- FUNCTION: CHAPERONE THAT MAY PLAY A ROLE IN THE CORRECT ASSEMBLY,
 INCLUDING TWO-COMPONENT SIGNAL-TRANSDUCTION SYSTEMS. MIGHT CO-
 OPERATE WITH DnaK TO ACTIVATE THE Rcs TWO-COMPONENT SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Type III membrane protein.
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D10483; ; NOT_ANNOTATED_CDS.
 DR EMBL: AE000116; AAC73166.1; -
 DR HSSP: P25685; LHDJ.
 DR EcoGene: EGI1570; djla.
 DR InterPro: IPR001623; DnaJ_N.
 DR Pfam: PF00226; DnaJ; 1.
 DR SMART: SM00271; DnaJ; 1.
 DR PROSITE: PS00636; DnaJ_1; FALSE_NEG.
 DR PROSITE: PS00076; DnaJ_2; 1.
 KW Chaperone; Transmembrane; Inner membrane; Complete proteome.
 FT DOMAIN 1 6 PERIPLASMIC (POTENTIAL).
 FT TRANSMEM 7 27 POTENTIAL.
 FT DOMAIN 28 271 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 205 271 J-DOMAIN.
 FT MUTAGEN 15 15 L->R: LOSS OF ACTIVATION OF RCS.
 FT MUTAGEN 16 16 M->R: ONLY PARTIAL ACTIVATION OF RCS.
 FT MUTAGEN 23 233 H->Q: LOSS OF ACTIVITY.
 SQ SEQUENCE 271 AA; 30579 MW; 80A0FC28F6D470DF CRC64;

Query Match 2.8%; Score 7; DB 1; Length 271;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 AKOQAE 80
 |||||
 DB 250 AKOQAE 256

RESULT 14
 DAAA_BACLI STANDARD; PRT: 283 AA.
 AC P54692;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate
 aminotransferase) (D-amino acid aminotransferase) (D-amino acid
 transaminase) (DAAT).
 GN DAT.
 OS Bacillus licheniformis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1402;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 10716;
 RX MEDLINE-97157074; PubMed-9003455;
 RA Taylor P.P., Fotheringham I.G.;
 RT "Nucleotide sequence of the Bacillus licheniformis ATCC 10716 dat
 gene and comparison of the predicted amino acid sequence with those
 of other bacterial species."
 RL Biochim. Biophys. Acta 1350:38-40(1997).
 CC -1- CATALYTIC ACTIVITY: D-ALANINE + 2-OXOGLUTARATE = PYRUVATE +
 D-GLUTAMATE (ACTS ON THE D-ISOMERS OF LEUCINE, ASPARTATE,
 GLUTAMATE, AMINO BUTYRATE, NORVALINE AND ASPARAGINE).
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 AMINOTRANSFERASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

```

CC EMBL; U26947; AAB50428.1; -
DR HSSP; P19338; IDAA.
DR InterPro; IPR001544; Aminotran_4.
DR Pfam; PF01063; aminotran_4; 1.
DR ProDom; PD001961; Aminotran_4; 1.
DR PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW PROSITE; Aminotransferase; Pyridoxal phosphate.
FT BINDING 144 144 PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ SEQUENCE 283 AA; 31396 MW; 2CA8FA604FE9D7 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 283;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 KOKAQEA 81
DB 155 KOKAQEA 161
|||||

RESULT 15
NU2M_ALBCO STANDARD; PRT; 307 AA.
AC P48902;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN ND2.
OS Albinaria coerulesa (Land snail).
OC Mitochondrion.
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Clausiliidae; Altopinae; Albinaria.
OX NCBI_TaxID=42349;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120351; PubMed=7498775;
RA Hatzoglou E., Rodakis G.C., Lecanidou R.;
RT "Complete sequence and gene organization of the mitochondrial genome of the land snail Albinaria coerulesa.";
RL Genetics 140:1353-1366(1995).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone -> NAD(+) + ubiquinol.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial inner membrane.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; X83390; CAA58306.1; -
DR InterPro; IPR001750; Oxidored_g1.
DR Pfam; PF00361; oxidored_g1; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 307 AA; 34865 MW; CF27C405278CB92D CRC64;

Query Match 2.8%; Score 7; DB 1; Length 307;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 SISLSLL 186
DB 218 SISLSLL 224
|||||

RESULT 16
NU2M_MYXGL STANDARD; PRT; 348 AA.
AC Q21078; O63918;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN MTND2 OR ND2 OR NAD2 OR NADH2.
OS Myxine glutinosa (Atlantic hagfish).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
OC Myxiniidae; Myxiniinae; Myxine.
OX NCBI_TaxID=7769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97398704; PubMed=9254918;
RA Delarbre C., Barriol V., Tillier S., Janvier P., Gachelin G.;
RT "The main features of the craniate mitochondrial DNA between the ND1 and the COI genes were established in the common ancestor with the lancelet.";
RL Mol. Biol. Evol. 14:807-813(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98210228; PubMed=9541532;
RA Rasmussen A.S., Janke A., Arnason U.;
RT "The mitochondrial DNA molecule of the hagfish (Myxine glutinosa) and vertebrate phylogeny.";
RL J. Mol. Evol. 46:382-388(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Delarbre C., Rasmussen A.S., Arnason U., Gachelin G.;
RT "Complete sequence of the mitochondrial DNA of Myxine glutinosa.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone -> NAD(+) + ubiquinol.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial inner membrane.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; Y09527; CAA70717.1; -
DR EMBL; Y15187; CAA75486.1; -
DR EMBL; AJ404477; CAC20650.1; -
DR InterPro; IPR003917; NADHUb_oxdrdclse2.
DR InterPro; IPR001750; Oxidored_g1.
DR Pfam; PF00361; oxidored_g1; 1.
DR PRINTS; PR01436; NADHHDGNASE2.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
FT CONFLICT 205 205 I -> V (IN REF. 2).
FT CONFLICT 315 315 S -> C (IN REF. 2).
SQ SEQUENCE 348 AA; 38788 MW; DD123DADDICF61EA CRC64;

Query Match 2.8%; Score 7; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 ISNSLIL 110
DB 269 ISNSLIL 275
|||||

RESULT 17
IADA_ECOLI STANDARD; PRT; 390 AA.
ID IADA_ECOLI
AC P39377;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Isoaspartyl dipeptidase (EC 3.4.19.-).
 GN IADA OR B4328.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC STRAIN-K12 / W3110; PubMed=7876157;
 RA MEDLINE=95181377; PubMed=7876157;
 RT Gary J.D., Clarke S.;
 RT "Purification and characterization of an isoaspartyl dipeptidase from
 RT Escherichia coli.";
 RL J. Biol. Chem. 270:4076-4087(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=9534362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 CC -!- FUNCTION: CATALYZES THE HYDROLYTIC CLEAVAGE OF A SUBSET OF L-
 CC ISOASPARTYL (L-BETA-ASPARTYL) DIPEPTIDES. USED TO DEGRADE PROTEINS
 CC DAMAGED BY L-ISOASPARTYL RESIDUES FORMATION.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: SOME, IN THE N-TERMINAL, TO DIHYDROOROTASE.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M38.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U15029; AAC43299.1; .
 DR EMBL; U14003; AAA97224.1; .
 DR EMBL; AE000503; AAC77284.1; .
 DR MEROPS; M38.001; .
 DR Ecogene; EG12567; iada.
 KW Hydrolase; Metalloprotease; Complete proteome.
 SQ SEQUENCE 390 AA; 41084 MW; 9CEEC838381545B5 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 390;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ISKLPT 95
 DB 223 ISKLPT 229
 |||||

RESULT 18
 ID YV4Q_CAEEL STANDARD; PRY; 413 AA.
 AC 045435;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 49.0 kDa protein F32B6.9 in chromosome IV.
 GN F32B6.9.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabditidae; Peleoderinae; Caenorhabditis.
 CC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;

RA Basham V.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE BESTROPHIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z81074; CAB03043.1; .
 DR WormPep; F32B6.9; CE09864.
 DR InterPro; IPR000615; Worm.fam.8.
 DR Pfam; PF01062; Worm.family_8; 1.
 DR ProDom; PD002802; Worm_fam_8; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 413 AA; 48965 MW; A66E59A83C78790B CRC64;

Query Match 2.8%; Score 7; DB 1; Length 413;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 PIDCKG 122
 DB 384 PIDCKG 390
 |||||

RESULT 19
 ID GCH2_CHLMU STANDARD; PRY; 424 AA.
 AC Q9PLJ5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Riboflavin biosynthesis protein ribA [Includes: GTP cyclohydrolase II
 DE (EC 3.5.4.25); 3,4-dihydroxy-2-butanone 4-phosphate synthase (DHBP
 DE synthase)].
 DE GN RIBAB OR TC0104.
 OS Chlamydia muridarum.
 CC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 CC NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MoPn / N199;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy K., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -!- CATALYTIC ACTIVITY: GTP + 3 H(2)O = formate + 2,5-diamino-6-
 CC hydroxy-4-(5-phosphoribosylamino)pyrimidine + diphosphate.
 CC -!- PATHWAY: RIBOFLAVIN BIOSYNTHESIS.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE DHBP
 CC SYNTHASE FAMILY.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GTP
 CC CYCLOHYDROLASE II FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF002278; AAF38984.1; .

```

DR TIGR: TC0104; ..
DR InterPro: IPR000422; DHPB_synthase.
DR InterPro: IPR000926; GTP_cyclohydro2.
DR Pfam: PF00926; DHPB_synthase; 1.
DR Pfam: PF00925; GTP_cyclohydro2; 1.
DR ProDom: PD001034; DHPB_synthase; 1.
KW Multifunctional enzyme; Riboflavin biosynthesis; Hydrolase;
KW Complete proteome.
FT DOMAIN 1 206 DHPB SYNTHASE.
FT DOMAIN 207 424 GTP CYCLOHYDROLASE II.
SQ SEQUENCE 424 AA: 47066 MW: 0298128162BAC016 CRC64;

Query Match 2.88; Score 7; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 SDPTSIS 182
Db 117 SDPTSIS 123

RESULT 20
YQ53_BACAN STANDARD: PRT: 482 AA.
AC Q9RMX9;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical trans-acting regulator pX02-53.
GS PX02-53.
OS Bacillus anthracis.
OC Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA Kumano S., Manter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ATXA/ACPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF188935; AAP13658.1;
SQ SEQUENCE 482 AA: 57210 MW: 8473BD1FDCA73315 CRC64;

Query Match 2.88; Score 7; DB 1; Length 482;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 EKLKVDL 62
Db 301 EKLKVDL 307

RESULT 21
F1S1_ECOLI STANDARD: PRT: 588 AA.
AC P04286;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE DE Peptidoglycan synthetase ftsI precursor (Penicillin-binding protein 3)
DE (PBP-3).
GN FTSI OR PBPB OR B0084 OR 20094 OR ECS0088.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=83296957; PubMed=6350821;
RA Nakamura M., Maruyama I.N., Soma M., Kato J., Suzuki H., Horota Y.;
RT "On the process of cellular division in Escherichia coli: nucleotide
RT sequence of the gene for penicillin-binding protein 3.";
RL Mol. Gen. Genet. 191:1-9(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posal G., Hackett J., Kline S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohlsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=93077472; PubMed=1447153;
RA Ueki M., Wachi M., Jung H.K., Ishino F., Matsuhashi M.;
RT "Escherichia coli mrar gene involved in cell growth and division.";
RL J. Bacteriol. 174:7841-7843(1992).
RN [7]
RP SEQUENCE OF 1-41 FROM N.A.
RX MEDLINE=93077455; PubMed=1332942;
RA Guzman L.M., Barondess J.J., Beckwith J.;
RT "ftsL, an essential cytoplasmic membrane protein involved in cell
RT division in Escherichia coli.";
RL J. Bacteriol. 174:7716-7728(1992).
RN [8]
RP SEQUENCE OF 550-588 FROM N.A.

```


RA MEDLINE-90328986; PubMed-2198024;
RA Michaud C., Parquet C., Flouret B., Blanot D., van Heijenoort J.,
RT "Revised interpretation of the sequence containing the murE gene
RT encoding the UDP-N-acetylmuramyl tripeptide synthetase of Escherichia
RT coli";
RL Biochem. J. 269:277-280(1990).
RN [9]
RN MUTAGENESIS OF SER-307.
RX MEDLINE-86117937; PubMed-3911028;
RA Houbi-Herlin N., Hara H., Inouye M., Hirota Y.,
RT "Binding of penicillin to thiol-penicillin-binding protein 3 of
RT Escherichia coli: identification of its active site";
RL Mol. Gen. Genet. 201:499-504(1985).
RN [10]
RN PROCESSING.
RX MEDLINE-90036670; PubMed-2681146;
RA Nagasawa H., Sakagami Y., Suzuki H., Suzuki H., Hirota Y.,
RT "Determination of the cleavage site involved in C-terminal processing
RT of penicillin-binding protein 3 of Escherichia coli";
RL J. Bacteriol. 171:5890-5893(1989).
RN [11]
RN MUTANT PBPER1.
RX MEDLINE-89008105; PubMed-3049550;
RA Taschner P.E., Ypenburg N., Spratt B.G., Woldringh C.L.,
RT "An amino acid substitution in penicillin-binding protein 3 creates
RT pointed polar caps in Escherichia coli";
RL J. Bacteriol. 170:4828-4837(1988).
RN [12]
RN TOPOLOGY.
RX MEDLINE-90014188; PubMed-2677607;
RA Bowler L.D., Spratt B.G.,
RT "Membrane topology of penicillin-binding protein 3 of Escherichia
RT coli";
RL Mol. Microbiol. 3:1277-1286(1989).
CC -!- FUNCTION: CELL WALL FORMATION. ESSENTIAL FOR THE FORMATION OF A
CC SEPTUM OF THE MUREIN SACCCULUS. SYNTHESIS OF CROSS-LINKED
CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES.
CC -!- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE. THE BULK OF THE MOLECULE,
CC EXCEPT FOR THE N-TERMINAL MEMBRANE ANCHOR REGION, PROTRUDES INTO
CC THE PERIPLASMIC SPACE. WHERE IT ACTS ON MUREIN.
CC -!- DOMAIN: HAS AN N-TERMINAL PENICILLIN INSENSITIVE TRANSGLYCOSYLASE
CC DOMAIN (FORMATION OF LINEAR GLYCAN STRANDS) & A CARBOXY-TERMINAL
CC PENICILLIN-SENSITIVE TRANSEPTIDASE DOMAIN (CROSS-LINKING OF THE
CC PEPTIDE SUBUNITS).
CC -!- MISCELLANEOUS: MUTANT PBPER1 WHICH WAS OBTAINED AFTER SELECTION
CC FOR INCREASED RESISTANCE TO CEPHALEXIN, CAUSES A CHANGE IN THE
CC SHAPE OF THE CELL. THE POLAR CAPS ARE POINTED.
CC -!- SIMILARITY: BELONGS TO THE TRANSEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K00137; AAA24300.1; -;
DR EMBL; X55034; CAA38861.1; -;
DR EMBL; D10483; BAA01349.1; -;
DR EMBL; AE000118; AAC73195.1; -;
DR EMBL; AE005185; AAC54388.1; -;
DR EMBL; AP002550; BAB33511.1; -;
DR EMBL; S49802; AAB24312.1; -;
DR EMBL; S49875; AAB24310.1; -;
DR EMBL; X55814; CAA39333.1; -;
DR PIR; A03419; ZPECP3.
DR PIR; B45278; B45278.
DR PIR; S40594; S40594.
DR Eogene; EG10341; tcsi.
DR InterPro; IPR001460; Transseptidase.
DR Pfam; PF00905; Transseptidase; 1.

KW Inner membrane; Peptidoglycan synthesis; Cell division; Cell wall;
KW Antibiotic resistance; Multifunctional enzyme; Cell shape;
KW Transmembrane; Complete proteome.
FT CHAIN 1 577 PEPTIDOGLYCAN SYNTHETASE FTSL.
FT PROPEP 578 588
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 39 POTENTIAL.
FT DOMAIN 40 577 PERIPLASMIC (POTENTIAL).
FT ACT_SITE 307 307 ACYLATED BY PENICILLIN.
FT VARIANT 361 361 N -> S (IN MUTANT PBPER1).
FT MUTAGEN 307 307 S->A,T: UNABLE TO BIND PENICILLIN.
FT MUTAGEN 307 307 S->C: STILL ABLE TO BIND PENICILLIN.
SQ SEQUENCE 588 AA: 63877 MW: C89A403D5980B2CD CRC64;
Query Match 2.8%; Score 7; DB 1; Length 588;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 63 GVLQKSS 69
| | | | | | | | | |
DB 354 GVLQKSS 360
RESULT 22
PLBI_YEAST
ID PLBI_YEAST STANDARD; PRT; 664 AA.
AC P39105;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lyso-phospholipase 1 precursor (EC 3.1.1.5) (Phospholipase B 1).
GN PLBI OR YMR008C OR YMR270.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE-94327513; PubMed-8051052;
RA Lee K.S., Patton J.L., Fido M., Hines L.K., Kohlwein S.P.,
RA Paltauf F., Henry S.A., Levin D.E.,
RT "The Saccharomyces cerevisiae PLBI gene encodes a protein required
RT for lyso-phospholipase and phospholipase B activity";
RL J. Biol. Chem. 269:19725-19730(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE RELEASE OF FATTY ACIDS FROM
CC LYSPHOSPHOLIPIDS.
CC -!- CATALYTIC ACTIVITY: 2-lyso-phosphatidylcholine + H(2)O =
CC glycerophosphocholine + a fatty acid anion.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: TO OTHER FUNGAL LYSPHOSPHOLIPASES AND TO YEAST
CC SPO1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L23089; AAA61611.1; -;
DR EMBL; Z48613; CAA88523.1; -;
DR PIR; A53647; A53647.
DR SGD; S0004610; PLBI.
DR InterPro; IPR002642; PLAC.
DR Pfam; PF01735; PLAC2_B; 1.

```

DR SMART: SM00022; PLAC; 1.
KW Lipid degradation: Hydrolase; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL
FT CHAIN 23 664 LYSOPHOSPHOLIPASE 1.
FT CARBOHYD 26 26 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 33 33 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 52 52 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 78 78 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 160 160 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 277 277 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 388 388 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 459 459 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 541 541 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 565 565 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 641 641 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CONFLICT 32 A -> S (IN REF. 1)
FT CONFLICT 494 E -> D (IN REF. 1)
FT CONFLICT 494 E -> D (IN REF. 1)
SQ SEQUENCE 664 AA: 71667 MW: E05A585E7AB73F34 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 664;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 GLNLSFP 128
Db 457 GLNLSFP 463
|||||

RESULT 23
YJ80_YEAST STANDARD; PRT; 588 AA.
ID YJ80_YEAST
AC P47147;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 80.2 kDa protein in CPA2-NNF1 intergenic region.
GN YJ8110W OR J2007
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramezani Rad M., Kirchthath U., Hollenberg C.P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koetter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.
CC -|- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z49610; CAA89640.1; -
CC SGD: S0003871; YJ8110W.
DR InterPro: IPR000340; DS_phosphatase.

```

```

DR SMART: SM000387; TYR_phosphatase.
DR InterPro: SM00012; PTPC_DSG; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
KW Hypothetical protein; Hydrolase.
FT ACT_SITE 397 397 BY SIMILARITY.
SQ SEQUENCE 688 AA: 80151 MW: A097F4B98B626C01 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 688;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 GKGINLS 126
Db 672 GKGINLS 678
|||||

RESULT 24
SIN4_YEAST STANDARD; PRT; 974 AA.
ID SIN4_YEAST
AC P32259;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Global transcriptional regulator SIN4.
GN SIN4 OR TSF3 OR BEL2 OR GAL22 OR SSF5 OR YNL236W OR N1135.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Chien S., West R.W. Jr., Johnson S.L., Gans H., Kruger B., Ma J.;
RL MEDLINE=93140781; PubMed=8423805;
RA Chien S., West R.W. Jr., Johnson S.L., Gans H., Kruger B., Ma J.;
RT "TSF3, a global regulatory protein that silences transcription of
RT yeast GAL genes, also mediates repression by alpha 2 repressor and is
RT identical to SIN4";
RL Mol. Cell. Biol. 12:4503-4514(1992);
RN [3]
RP SEQUENCE FROM N.A.
RA Harashima S., Mabuchi H., Ramash R., Hasebe M., Tanaka A., Oshima Y.;
RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA MEDLINE=97051546; PubMed=8896273;
RA Pandolfo B., de Antoni A., Laniranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
RT reading frames including a novel gene encoding a globin-like
RT domain.";
RL Yeast 12:1071-1076(1996).
CC -|- FUNCTION: GLOBAL REGULATORY PROTEIN THAT SILENCES TRANSCRIPTION
CC OF GAL AND MATING-TYPE GENES. NEGATIVE REGULATOR OF THE HO
CC (HOMOTHALLISM) GENE. MAY POTENTIATE TRANSCRIPTIONAL ACTIVATION
CC AND REPRESSION BY REGULATING THE ACTIVITY OF THE BASAL
CC TRANSCRIPTIONAL APPARATUS. SIN4 MUTATION ALTERS CHROMATIN
CC STRUCTURE.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC EMBL: M93050; AAA35044.1;
DR EMBL: X64516; CAA45819.1;
DR EMBL: D12918; BAA02302.1;
DR EMBL: 269381; CAA93362.1;
DR EMBL: 271512; CAA96140.1;
DR PIR: A44484; A44484.
DR PIR: S20132; S20132.
DR PIR: A48074; A48074.
DR PIR: S41805; S41805.
DR TRANSFAC: T01243.
DR SGD: S0005180; SIN4.
KW Transcription regulation; Activator; DNA-binding; Repressor;
KW Nuclear protein.
FT DOMAIN 63 81 SER/THR-RICH.
FT DOMAIN 889 893 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 974 AA; 111296 MW; 12ECF5B4CPC05A8E CRC64;

Query Match 2.8%; Score 7; DB 1; Length 974;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GLKISNS 107
DB 859 GLKISNS 865
|||||

RESULT 25
SIP3_YEAST
ID SIP3_YEAST STANDARD; PRT: 1229 AA.
AC P38717.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SIP3 protein.
GN SIP3 OR YNL257C OR N0844.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94173726; PubMed=8127709;
RA Lesage P., Yang X., Carlson M.;
RT "Analysis of the SIP3 protein identified in a two-hybrid screen for
RT interaction with the SNF1 protein kinase."
RL Nucleic Acids Res. 22:597-603(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FYJ679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueidener U., Beinbauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SU11
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae."
RL Yeast 13:849-860(1997).
CC -!- FUNCTION: INTERACTS WITH THE SNF1 PROTEIN KINASE.
CC -!- SIMILARITY: STRONG, TO YEAST YHR155W.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U03376; AAA17885.1;
DR EMBL: X96722; CAA65487.1;
DR EMBL: 271533; CAA96164.1;

```

```

DR PIR: S42391; S42391.
DR SGD: S0005201; SIP3.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PS00003; PH_DOMAIN; 1.
FT DOMAIN 309 423
SQ SEQUENCE 1229 AA; 142819 MW; 977995219282CD65 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 1229;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 IKASLDL 152
DB 190 IKASLDL 196
|||||

RESULT 26
SC25_YEAST
ID SC25_YEAST STANDARD; PRT: 1253 AA.
AC P14771.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SCD25 protein.
GN SCD25 OR SDC25 OR YLL016W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W303;
RX MEDLINE=91094833; PubMed=1986220;
RA Danak F., Boy-Marcotte E., le Roscouet D., Guilbaud R., Jacquet M.;
RT "The C-terminal part of a gene which contains a RAS-activating domain and
RT is a dispensable gene of Saccharomyces cerevisiae."
RL Mol. Cell. Biol. 11:202-212(1991).
RN [2]
RP SEQUENCE OF 668-1253 FROM N.A.
RC STRAIN=01136;
RX MEDLINE=99306677; PubMed=2545538;
RA Boy-Marcotte E., Danak F., Camonis J., Garreau H., Jacquet M.;
RT "The C-terminal part of a gene partially homologous to CDC 25 gene
RT suppresses the cdc25-5 mutation in Saccharomyces cerevisiae."
RL Gene 77:21-30(1989).
RN [3]
RP FUNCTION.
RX MEDLINE=90260633; PubMed=2188363;
RA Crechet J.B., Pouillet P., Mistou M.-Y., Parmeggiani A., Camonis J.,
RA Boy-Marcotte E., Danak F., Jacquet M.;
RT "Enhancement of the GDP-GTP exchange of RAS proteins by the carboxyl-
RT terminal domain of SCD25."
RL Science 248:866-868(1990).
RN [4]
RP FUNCTION.
RX MEDLINE=91156312; PubMed=2000228;
RA Rey I., Schweighoffer F., Barlat I., Camonis J., Boy-Marcotte E.,
RA Guilbaud R., Jacquet M., Tocque B.;
RT "The COOH-domain of the product of the Saccharomyces cerevisiae SCD25
RT gene elicits activation of p21-ras proteins in mammalian cells."
RL Oncogene 6:347-349(1991).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -!- MISCELLANEOUS: SUPPRESSES THE CDC25-5 MUTATION IN YEAST (RESTORES
CC CAMP LEVEL) AND HAS SIMILAR FUNCTIONS AS CDC25.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M26647; AAA16565.1; -
 CC PIR: S14177; S14177
 CC SGD: S0003939; SDC25.
 CC InterPro: IPR000651; RasGEFN.
 CC InterPro: IPR001895; RasGRP_CDC25.
 CC InterPro: IPR001452; SH3.
 CC Pfam: PF00617; RasGEF; 1.
 CC Pfam: PF00618; RasGEFN; 1.
 CC Pfam: PF00018; SH3; 1.
 CC SMART: SM00147; RasGEF; 1.
 CC SMART: SM00229; RasGEFN; 1.
 CC SMART: SM00326; SH3; 1.
 CC PROSITE: PS00720; GDS_CDC25; 1.
 CC PROSITE: PS50002; SH3; 1.
 CC Guanine-nucleotide releasing factor; Cell cycle; Cell division;
 KW SH3 domain. 26 98 SH3.
 FT DOMAIN 74 79 POLY-ASN.
 FT DOMAIN 434 437 POLY-ARG.
 FT VARIANT 584 590 DVVVKFI -> V (IN STRAIN OL136).
 SQ SEQUENCE 1253 AA: 144979 MW: 2DE2C9EC27E3E60D CRC64;

Query Match 2.8%; Score 7; DB 1; Length 1253;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 149 SLDLLTA 155
 |||||
 DB 254 SLDLLTA 260

RESULT 27

NI70_YEAST
 ID NI70_YEAST STANDARD; PRT: 1502 AA.
 AC P38181;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nucleoporin NUP170 (Nuclear pore protein NUP170).
 GN NUP170 OR YBL079W OR YBL0725.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=96076635; PubMed=7502586;
 RA Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;
 RT "Sequence analysis of a 78.6 kb segment of the left end of
 RT Saccharomyces cerevisiae chromosome II.";
 RL Yeast 11:1103-1112(1995).
 RN [2]
 RP SEQUENCE OF 1262-1502 FROM N.A.
 RC STRAIN=S288C;
 RA Contreras R., Fiers W., Logghe M., Molemans F.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION, AND SEQUENCE OF 117-133 AND 153-166.
 RX MEDLINE=96095775; PubMed=8522578;
 RA Aitchison J.D., Rout M.P., Marelli M., Blobel G., Wozniak R.W.;
 RT "Two novel related yeast nucleoporins Nup170p and Nup157p;
 RT complementation with the vertebrate homologue Nup155p and functional
 RT interactions with the yeast nuclear pore-membrane protein Pom152p.";
 RL J. Cell Biol. 131:1133-1148(1995).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.

CC -!- SIMILARITY: TO YEAST NUP157, AND SOME, TO MAMMALIAN NUP155.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X79489; CAAS6029.1; -
 DR EMBL: Z35840; CAA84900.1; -
 DR PIR: S45429; S45429.
 DR SGD: S0000175; NUP170.
 KW Nuclear protein; Transport.
 SQ SEQUENCE 1502 AA: 169474 MW: 3BEA65DAA2A5F99A CRC64;

Query Match 2.8%; Score 7; DB 1; Length 1502;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 206 STVSSLL 212
 |||||
 DB 723 STVSSLL 729

RESULT 28

GP21_RAT
 ID GP21_RAT STANDARD; PRT: 1886 AA.
 AC P11654;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Integral membrane glycoprotein gp210 precursor.
 GN GP210.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89291948; PubMed=2738089;
 RA Wozniak R.W., Bartnik E., Blobel G.;
 RT "Primary structure analysis of an integral membrane glycoprotein of
 RT the nuclear pore.";
 RL J. Cell Biol. 108:2083-2092(1989).
 RN [2]
 RP TOPOLOGY.
 RX MEDLINE=93107146; PubMed=1281815;
 RA Wozniak R.W., Blobel G.;
 RT "The single transmembrane segment of gp210 is sufficient for sorting
 RT to the pore membrane domain of the nuclear envelope.";
 RL J. Cell Biol. 119:1441-1449(1992).
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. NUCLEAR PORE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y00826; CAA68759.1; -
 DR PIR: S04921; S04921.
 DR InterPro: IPR003343; Big_2.
 DR Pfam: PF02368; Big_2; 1.
 KW Glycoprotein; Transmembrane; Nuclear protein; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1886
 FT DOMAIN 26 1805 INTEGRAL MEMBRANE GLYCOPROTEIN GP210.
 FT TRANSMEM 1806 1828 CISTERNAL SIDE (PROBABLE).
 FT

FT DOMAIN 1829 1886 CYTOPLASMIC (PROBABLE).
 FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 681 681 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1039 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1116 1116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1135 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1362 1362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1441 1441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1886 AA; 204158 MW; 6920B93C20A6C5D1 CRC64;
 Query Match 2.8%; Score 7; DB 1; Length 1886;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 180 SISLSLL 186
 DQ 689 SISLSLL 695
 RESULT 29
 SPCB_DROME STANDARD; PRT; 2291 AA.
 AC Q00963; Q9VX30;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Spectrin beta chain.
 GN BETA-SPEC OR SPEC-B OR CG5870.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE OF 1-800 FROM N.A.
 RX MEDLINE=9009037; PubMed=2677025;
 RA Byers T.J., Husain-Chisti A., Dubreuil R.R., Branton D.,
 RA Goldstein L.S.B.;
 RT "Sequence similarity of the amino-terminal domain of Drosophila beta
 RT spectrin to alpha actinin and dystrophin".
 RL J. Cell Biol. 109:1633-1641(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9235263; PubMed=1631106;
 RA Byers T.J., Branton D., Lue R., Winograd E., Branton D.;
 RT "The complete sequence of Drosophila beta-spectrin reveals
 RT supra-motifs comprising eight 106-residue segments".
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6187-6191(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Geliniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Sinden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster".
 Science 287:2185-2195(2000).
 [4]
 CHARACTERIZATION.
 MEDLINE=88059242; PubMed=3680372;
 Dubreuil R., Byers T.J., Branton D., Goldstein L.S.B., Kiehart D.P.;
 "Drosophila spectrin. I. Characterization of the purified protein".
 J. Cell Biol. 105:2095-2102(1987).
 [5]
 STRUCTURE BY NMR OF 2145-2262.
 MEDLINE=96164435; PubMed=8591029;
 Zhang P., Talluri S., Deng H., Branton D., Wagner G.;
 "Solution structure of the pleckstrin homology domain of Drosophila
 beta-spectrin".
 Structure 3:1185-1195(1995).
 CC -!- FUNCTION: SPECTRIN IS THE MAJOR CONSTITUENT OF THE CYTOSKELETAL
 CC NETWORK UNDERLYING THE ERYTHROCYTE PLASMA MEMBRANE. IT ASSOCIATES
 CC WITH BAND 4.1 AND ACTIN TO FORM THE CYTOSKELETAL SUPERSTRUCTURE OF
 CC THE ERYTHROCYTE PLASMA MEMBRANE. INTERACTS WITH CALMODULIN IN A
 CC CALCIUM-DEPENDENT MANNER.
 CC -!- SUBUNIT: NATIVE SPECTRIN MOLECULE IS A TETRAMER COMPOSED OF TWO
 CC ANTIPARALLEL HETERODIMERS JOINED HEAD TO HEAD SO THAT EACH END
 CC OF THE NATIVE MOLECULE INCLUDES THE C-TERMINUS OF THE ALPHA
 CC SUBUNIT AND THE N-TERMINUS OF THE BETA SUBUNIT.
 CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -!- SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL; M92288; AAA28399.1; -
 CC EMBL; AE003506; AAF48751.1; -
 CC PIR; A33657; A33657.
 CC PIR; A46147; A46147.
 CC PDB; 1DRO; 03-APR-96.
 CC FlyBase; FBgn0003471; Beta-Spec.
 CC InterPro; IPR001589; Actinin_act_bind.
 CC InterPro; IPR001715; Calponin_hom.
 CC InterPro; IPR001849; PH.
 CC InterPro; IPR002017; Spectrin.
 CC InterPro; IPR001605; Spectrin_ph.
 CC Pfam; PF00307; CH; 2.
 CC Pfam; PF00169; PH; 1.
 CC Pfam; PF00435; spectrin; 17.
 CC PRINTS; PR00683; SPECTRINPH.

QY 124 NLSPPVT 130
 |||||
 DB 773 NLSPPVT 779

RESULT 31
 TM2A_METMA STANDARD: PRT: 12 AA.
 ID TM2A_METMA
 AC P80652;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alternative tetrahydromethanopterin S-methyltransferase 28 kDa subunit
 DE (EC 2.1.1.86) (N5-methyltetrahydromethanopterin--coenzyme M
 methyltransferase 28 kDa subunit) (Fragment).
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea: Euryarchaeota: Methanosarcinales; Methanosarcinaceae;
 OC Methanosarcina
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=DSM 3647 / GOEL;
 RX MEDLINE=96370840; PubMed=8774736;
 RA Lienard T., Becher B., Marschall M., Bowien S., Gottschalk G.;
 RT "Sodium ion translocation by N5-methyltetrahydromethanopterin:
 coenzyme M methyltransferase from Methanosarcina mazei Gol
 reconstituted in ether lipid liposomes.";
 RL Eur. J. Biochem. 239:857-864(1996).
 CC -1- FUNCTION: THIS ENZYME COMPLEX CATALYZES AN INTERMEDIATE STEP IN
 METHANOGENESIS, THE FORMATION OF METHYL-COENZYME M AND
 TETRAHYDROMETHANOPTERIN FROM COENZYME M AND N5-METHYL-
 TETRAHYDROMETHANOPTERIN.
 CC -1- CATALYTIC ACTIVITY: 5-methyl-5,6,7,8-tetrahydromethanopterin + 2-
 mercaptoethanesulfonate -> 5,6,7,8-tetrahydromethanopterin + 2-
 (methylthio)ethanesulfonate.
 CC -1- SUBUNIT: COMPOSED OF SIX DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 KW Transferase; Methyltransferase; Transmembrane; Methanogenesis.
 FT NON_TER 12 12
 FT SEQUENCE 12 AA: 1321 MW; 6DE4A576623D76B CRC64;

Query Match 2.4%; Score 6; DB 1; Length 12;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 KLEPVL 40
 |||||
 DB 3 KLEPVL 8

RESULT 32
 CECB_ANTPE STANDARD: PRT: 35 AA.
 ID CECB_ANTPE
 AC P01509;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Cecropin B.
 OS Antheraea pernyi (Chinese oak silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Bombycoidea; Saturniidae; Saturniinae; Antheraea.
 OX NCBI_TaxID=7119;
 RN [1]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=83053368; PubMed=6754375;
 RA Qu X.-M., Steiner H., Engstroem A., Bennich H., Boman H.G.;
 RT "Insect immunity: Isolation and structure of cecropins B and D from
 pupae of the Chinese oak silk moth, Antheraea pernyi.";
 RL Eur. J. Biochem. 127:219-224(1982).
 RN [2]
 RP SEQUENCE.

RX MEDLINE=88108273; PubMed=2962676;
 RA Craig A.G., Engstrom A., Bennich H., Kamensky I.;
 RT "Plasma desorption mass spectrometry coupled with conventional
 peptide sequencing techniques.";
 RL Biomed. Environ. Mass Spectrom. 14:669-673(1987).
 CC -1- FUNCTION: CECROPINS HAVE LYTIC AND ANTIBACTERIAL ACTIVITY AGAINST
 SEVERAL GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA.
 CC -1- SIMILARITY: BELONGS TO THE CECROPIN FAMILY.
 DR PIR: A01771; CKA0BP.
 DR PIR: A54725; A54725.
 DR InterPro: IPR000875; Cecropin.
 DR InterPro: IPR003254; IIP_Cecropin.
 DR Pfam: PF00272; Cecropin; 1.
 DR ProDom: PD003996; IIP_Cecropin; 1.
 DR PROSITE: PS00268; CECROPIN; 1.
 KW Insect immunity; Antibiotic; Hemolymph; Amidation; Multigene family.
 FT MOD_RES 35 35
 FT AMIDATION.
 SQ SEQUENCE 35 AA: 3818 MW; 7AFCBB0A10E16313 CRC64;

Query Match 2.4%; Score 6; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 VAVLGE 173
 |||||
 DB 26 VAVLGE 31

RESULT 33
 TXC9_CUPSA STANDARD: PRT: 68 AA.
 ID TXC9_CUPSA
 AC P58604;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE Toxin CSTX-9
 DE Toxin CSTX-9
 OS Cupienius salei (Wandering spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Entelegynae; Lycosoidea; Ctenidae; Cupienius.
 OX NCBI_TaxID=6928;
 RN [1]
 RP SEQUENCE, DISULFIDE BONDS, MASS SPECTROMETRY, AND CHARACTERIZATION.
 RC TISSUE=Venom;
 RX MEDLINE=21544892; PubMed=11693332;
 RA Schaller J., Kaempfer U., Schuerch S., Kuhn-Nentwig L., Haeblerli S.,
 RA Nentwig W.;
 RT "CSTX-9, a toxic peptide from the spider Cupienius salei: amino acid
 sequence, disulphide bridge pattern and comparison with other spider
 toxins containing the cystine knot structure.";
 RL Cell. Mol. Life Sci. 58:1539-1545(2001).
 CC -1- FUNCTION: THIS TOXIN CAUSES PARALYSIS IN DROSOPHILA WITH AN LD50
 VALUE OF 3.12 PMOL/MG.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Produced by the venomous gland.
 CC -1- MASS SPECTROMETRY: MW=7529.75; MW_ERR=0.32; METHOD=Electrospray.
 KW Toxin; Neurotoxin.
 FT DISULFID 6 21
 FT DISULFID 13 30
 FT DISULFID 20 48
 FT DISULFID 32 46
 SQ SEQUENCE 68 AA: 7539 MW; 5BB526DCA359F9F4 CRC64;

Query Match 2.4%; Score 6; DB 1; Length 68;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 SLLQK 214
 |||||
 DB 50 SLLQK 55

```

RESULT 34
RP5M_THIFE STANDARD; PRT; 78 AA.
AC P24694;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Probable sigma(54) modulation protein (ORF3) (Fragment).
OS Thiobacillus ferrooxidans.
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.
OX NCBI_TaxID=920;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC_33020;
RA MEDLINE=90330545; PubMed=2198257;
RA Berger D.K., Woods D.R., Kawlings D.E.;
RT "Complementation of Escherichia coli sigma 54 (NtrA)-dependent
RT formate hydrogenlyase activity by a cloned Thiobacillus ferrooxidans
RT ntrA gene.";
RL J. Bacteriol. 172:4399-4406(1990).
CC -!- FUNCTION: PROBABLY MODULATES THE ACTIVITY OF SIGMA(54) (RPN).
CC -!- SIMILARITY: BELONGS TO THE SIGMA(54) MODULATION PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMFL; M58480; AAA27380.1;
DR PIR; C37761;
DR InterPro: IPR003489; Ribosomal_S30.
DR Pfam: PF02482; Ribosomal_S30; 1.
DR NON_TER 78
FT SEQUENCE 78 AA; 8751 MW; F39C7C4944E413EB CRC64;

Query Match 2.4%; Score 6; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LSNVVD 34
DB 49 LSNVVD 54

RESULT 35
SSS2_SCYCA STANDARD; PRT; 80 AA.
AC P1020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Spermatid-specific protein S2 (Basic nuclear protein S2).
OS Scyllorhinus canicula (Spotted dogfish) (Spotted catshark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC Scyllorhinidae; Scyllorhinus.
OX NCBI_TaxID=7830;
RN [1]
RP SEQUENCE
RC MEDLINE=89170733; PubMed=2924768;
RA Chauviere M., Martinage A., Briand G., Sautiere P., Chevallier P.;
RT "Nuclear basic protein transition during sperm differentiation.
RT Primary structure of the spermatid-specific protein S2 from the
RT dog-fish Scyllorhinus caniculus.";
RL Eur. J. Biochem. 180:329-335(1989).
CC -!- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION; HISTONES
CC ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES
CC REPLACED BY PROTAMINES IN LATE SPERMATIDS.
CC -!- SUBCELLULAR LOCATION: Nuclear.

```

```

CC -!- MISCELLANEOUS: N-TERMINAL HALF IS HIGHLY BASIC, WHILE C-TERMINAL
CC PART IS ACID.
CC -!- SIMILARITY: WITH PROTEIN S1 (70% STRUCTURAL SIMILARITY).
CC PIR: S03560;
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Nuclear protein.
SQ SEQUENCE 80 AA; 9733 MW; 593E18EF33452409 CRC64;

Query Match 2.4%; Score 6; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VDNTIK 52
DB 55 VDNTIK 60

RESULT 36
Y039_SYNPF STANDARD; PRT; 81 AA.
ID Y039_SYNPF
AC P27368;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 9.2 kDa protein in cyst-cysr intergenic region (ORF 81).
OS Synecococcus sp. (Strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=91210162; PubMed=1708375;
RA Laudenbach D.E., Grossman A.R.;
RT "Characterization and mutagenesis of sulfur-regulated genes in a
RT cyanobacterium: evidence for function in sulfate transport.";
RL J. Bacteriol. 173:2739-2750(1991).
CC -!- FUNCTION: MAY HAVE A REGULATORY FUNCTION.
CC -!- SIMILARITY: TO SYNECHOCYSTIS PCC 6803 SSR2439.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M65247; AAA73045.1;
DR PIR; D43670; D43670.
KW Sulfate transport; Hypothetical protein.
SQ SEQUENCE 81 AA; 9251 MW; 2F4EEC6E3E82C146 CRC64;

Query Match 2.4%; Score 6; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LTGTSE 19
DB 52 LTGTSE 57

RESULT 37
RS20_MYCPN STANDARD; PRT; 87 AA.
ID RS20_MYCPN
AC P75237;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S20.
OS Mycoplasma pneumoniae.
ON REST OR MPN541 OR MP301.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

```


OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RA MEDLINE-97105885; PubMed-8948633;
RX Hammelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -I- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -I- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000028; AAB95949.1;
DR InterPro: IPR002583; Ribosomal_S20p.
DR Pfam: PF01649; Ribosomal_S20p.1
DR ProDom: PD004231; Ribosomal_S20p.1
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 57 AA; 9970 MW; 4CAB26530C6B1D0C CRC64;

Query Match 2.4%; Score 6; DB 1; Length 87;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 LDNLGN 27
|
DB 41 LDNLGN 46

RESULT 38
ICBR_HUMAN
ID ICBR_HUMAN STANDARD; PRT; 90 AA.
AC P57730;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Caspase-1 inhibitor iceberg.
GN ICEBERG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20503847; PubMed-11051551;
RA Humke E.W., Shriver S.K., Starovasnik M.A., Fairbrother W.J.,
RA Dixit V.M.;
RT "ICEBERG: a novel inhibitor of interleukin-1beta generation.";
RL Cell 103:99-111(2000).
CC -I- FUNCTION: INHIBITS GENERATION OF IL-1BETA BY INTERACTING WITH
CC CASPASE-1 AND PREVENTING ITS ASSOCIATION WITH R12.
CC -I- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN THE HEART AND PLACENTA.
CC -I- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF208005; AAG23528.1;
DR

DR MIM: 605354;
DR InterPro: IPR001315; CARD.
DR Pfam: PF00619; CARD.1.
DR SMART: SM00114; CARD.1.
DR PROSITE: PSS0209; CARD.1.
KW Thiol protease inhibitor.
FT DOMAIN 1 90 CARD.
SQ SEQUENCE 90 AA; 10138 MW; 7E74C74263367101 CRC64;

Query Match 2.4%; Score 6; DB 1; Length 90;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 RIFIHS 226
|
DB 11 RIFIHS 16

RESULT 39
YP98_CAEEL
ID YP98_CAEEL STANDARD; PRT; 91 AA.
AC Q09246;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 10.2 kDa protein C28H8.8 in chromosome III.
GN C28H8.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Miller N., Waterston R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U20861; AAA62296.1;
DR WormPep: C28H8.8; CE01830.
DR InterPro: IPR002485; DUF13.
DR Pfam: PF01482; W0F13.1.
DR ProDom: PD002726; DUF13.1.
KW Hypothetical protein.
SQ SEQUENCE 91 AA; 10199 MW; HD4D391C6D1B1DAD CRC64;

Query Match 2.4%; Score 6; DB 1; Length 91;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 LDVKA 115
|
DB 33 LDVKA 38

RESULT 40
Y249_ARCFU
ID Y249_ARCFU STANDARD; PRT; 92 AA.
AC Q29990;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0249.
GN AF0249.

```

OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001088; AAB50984.1;
DR TIGR: AF0249;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 92 AA: 10460 MW: 718C209B54679EDF CRC64;

Query Match 2.4% Score 6; DR 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 EKLKVD 61
Db 8 EKLKVD 13

RESULT 41
MYLE_HUMAN
ID MYLE_HUMAN STANDARD; PRT: 95 AA.
AC O93424;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MYLE protein (Dexamethasone-induced protein).
GN MYLE OR DEXI
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Edgar A.J.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98049343; PubMed=9389475;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gasterland T., Gissi C., King B., Kochiwa H.,
RA Kuenli P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Hoffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mezzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection.
RL Nature 409:685-690(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF152470; AAB34163.2;
DR EMBL: AK007644; BAB25159.1; ALT_INIT.
DR MGD: MGI:1926236; Myle.

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF108145; AAC83382.2;
DR EMBL: BC001083; AAH01083.1; ALT_INIT.
SQ SEQUENCE 95 AA: 10429 MW: F9F98DE2572D3F83 CRC64;

Query Match 2.4% Score 6; DR 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 60 VDLGVL 65
Db 68 VDLGVL 73

RESULT 42
MYLE_MOUSE
ID MYLE_MOUSE STANDARD; PRT: 95 AA.
AC Q9MU07;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MYLE protein (Dexamethasone-induced protein).
GN MYLE OR DEXI
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Edgar A.J.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Pancreas;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gasterland T., Gissi C., King B., Kochiwa H.,
RA Kuenli P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Hoffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mezzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection.
RL Nature 409:685-690(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF152470; AAB34163.2;
DR EMBL: AK007644; BAB25159.1; ALT_INIT.
DR MGD: MGI:1926236; Myle.

```

SQ SEQUENCE 95 AA; 10402 MW; 4F293AD7F2326B14 CRC64;

Query Match 2.4%; Score 6; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 VDLGVL 65
Db 68 VDLGVL 73

RESULT 43
FSPM_LYCES STANDARD; PRT; 96 AA.
AC P14903;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Fruit-specific protein.
GN 2A11.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids 1; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, UC828; TISSUE=Fruit;
RX MEDLINE=91370848; PubMed=2491680;
RA Pear J.R., Ridge N.P., Rasmussen R., Rose R.E., Houck C.M.;
RT Isolation and characterization of a fruit-specific cDNA and the
RL corresponding genomic clone from tomato.*;
RL Plant Mol. Biol. 13:639-651(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, VENT CHERRY;
RA van Haaren M.J.J., Houck C.M.;
RL Submitted (XX-1990) to the EMBL/GenBank/DBJ databases.
CC -!- TISSUE SPECIFICITY: FRUIT-SPECIFIC.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SULFUR-RICH (MET AND CYS RICH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13743; CAA32007.1;
DR EMBL; M21775; AAA34129.1;
DR EMBL; M21776; AAA34130.1;
DR EMBL; M87659; AAA34165.1;
DR PIR; S07603; S07603.
SQ SEQUENCE 96 AA; 10699 MW; F6B2BF5D0A955FA6 CRC64;

Query Match 2.4%; Score 6; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 KLLPTN 96
Db 49 KLLPTN 54

RESULT 44
LTUB_CHLTR STANDARD; PRT; 97 AA.
ID LTUB_CHLTR
AC Q46404;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Late transcription unit B protein.
GN LTUB or CT080.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LGV434;
RX MEDLINE=95362647; PubMed=7543468;
RA Fahr M.J., Douglas A.L., Xia W., Hatch T.P.;
RT Characterization of late gene promoters of Chlamydia trachomatis.*;
RL J. Bacteriol. 177:4252-4260(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.I., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.*;
RL Science 282:754-759(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L40838; AAA75630.1;
DR EMBL; AF001282; AAC67671.1;
KW Complete proteome.
SQ SEQUENCE 97 AA; 11323 MW; CBE535A00351055D CRC64;

Query Match 2.4%; Score 6; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 SLIDKH 189
Db 60 SLIDKH 65

RESULT 45
YAFN_ECOLI STANDARD; PRT; 97 AA.
ID YAFN_ECOLI
AC Q47156; 1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yafn.
GN YAFN OR B0232.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Ohmori H.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SCHRAMM S., DUNCAN M., ALLEN E., ARAUJO R., APARICIO A., CHUNG E.,
 RA DAVIS K., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
 RA LASHKARI D., LEW H., LIN D., NAMATH A., OEFNER P., ROBERTS D.,
 RA DAVIS R.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D38582; BAA07594.1; --
 DR EMBL: AEO00131; AAC73336.1; --
 DR EMBL: U70214; AAB08652.1; --
 DR EcoGene; EG13151; yafN.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 97 AA: 11234 MW: C49811AE1E90A659 CRC64;

 Query Match 2.4%; Score 6; DB 1; Length 97;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 166 QPVAVL 171
 I I I I I I
 DB 26 QPVAVL 31

 RESULT 46
 C552_HYDTH STANDARD: PRT; 98 AA.
 ID C552_HYDTH 14, Created
 AC P15452;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Cytochrome c-552 precursor (C552).
 OS Hydrogenobacter thermophilus.
 OC Bacteria; Aquificales; Aquificaceae; Hydrogenobacter.
 OX NCBI_TaxID=940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TK-6;
 RX MEDLINE=91249816; PubMed=1645652;
 RA Sanbongi Y., Yang J.H., Igarashi Y., Kodama T.;
 RT "Cloning, nucleotide sequence and expression of the cytochrome c-552
 RT gene from Hydrogenobacter thermophilus.";
 RL Eur. J. Biochem. 198;7-12(1991).
 RN [2]
 RP SEQUENCE OF 19-98.
 RC STRAIN=TK-6;
 RX MEDLINE=89123087; PubMed=2536669;
 RA Sanbongi Y., Ishii M., Igarashi Y., Kodama T.;
 RT "Amino acid sequence of cytochrome c-552 from a thermophilic
 RT hydrogen-oxidizing bacterium, Hydrogenobacter thermophilus.";
 RL J. Bacteriol. 171:65-69(1989).
 RN [3]
 RP THERMOSTABILITY.
 RX MEDLINE=9012833; PubMed=2556725;
 RA Sanbongi Y., Igarashi Y., Kodama T.;
 RT "Thermostability of cytochrome c-552 from the thermophilic hydrogen-
 RT oxidizing bacterium Hydrogenobacter thermophilus.";
 RL Biochemistry 28:9574-9578(1989).
 RN [4]
 RP STRUCTURE BY NMR.
 RC STRAIN=TK-6;
 RX MEDLINE=98322065; PubMed=9657676;

RA Hasegawa J., Yoshida T., Yamazaki T., Sambongi Y., Yu Y., Igarashi Y.,
 RA Kodama T., Yamazaki K., Kyogoku Y., Kobayashi Y.;
 RT "Solution structure of thermostable cytochrome c-552 from
 RT Hydrogenobacter thermophilus determined by 1H-NMR spectroscopy.";
 RL Biochemistry 37:9641-9649(1998).
 CC -1- FUNCTION: REACTS WITH HYDROGENASE.
 CC -1- PTM: BINDS ONE HEME GROUP PER MOLECULE.
 CC -1- SIMILARITY: 56% WITH P.AEKUGINOSA C551.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X57735; CAA40902.1; --
 DR EMBL: A32226; A32226.
 DR PIR: S32485; S32485.
 DR PDB: 1AYG; 13-JAN-99.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR003088; Cyt_C1.
 DR InterPro; IPR002324; Cyt_C1D.
 DR Pfam; PF00034; cytochrome-c; 1.
 DR PRINTS; PRO0606; CYTOCHROME_C1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW Electron transport; Heme; Signal; 3D-structure.
 FT SIGNAL 1 18 CYTOCHROME C-552.
 FT CHAIN 19 98 HEME (COVALENT).
 FT BINDING 28 28 HEME (COVALENT).
 FT BINDING 31 31 HEME (COVALENT).
 FT METAL 32 32 IRON (HEME AXIAL LIGAND).
 FT METAL 77 77 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 98 AA: 10431 MW: F49713D829DDE927 CRC64;

 Query Match 2.4%; Score 6; DB 1; Length 98;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 72 QLAKOK 77
 I I I I I I
 DB 21 QLAKOK 26

 RESULT 47
 RK24_GUITH STANDARD: PRT; 101 AA.
 ID RK24_GUITH 37, Created
 AC O46905;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chloroplast 50S ribosomal protein L24.
 GN RPL24.
 OS Guillardia theta (Cryptomonas phi).
 OC Chlorophyta.
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
 OX NCBI_TaxID=55529;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97283757; PubMed=9137835;
 RA Wang S.L., Liu X.-Q., Douglas S.E.;
 RT "The large ribosomal protein gene cluster of a cryptomonad plastid:
 RT gene organization, sequence and evolutionary implications.";
 RL Biochem. Mol. Biol. Int. 41:1035-1044(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99128221; PubMed=9929392;
 RA Douglas S.E., Penny S.L.;
 RT "The plastid genome of the cryptophyte alga, Guillardia theta:
 RT complete sequence and conserved synteny groups confirm its common
 RT ancestry with red algae.";

RL J. Mol. Evol. 48:236-244(1999).
 CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF041468; AAC35714.1;
 DR InterPro: IPR003256; Ribosomal_L24_NusG.
 DR InterPro: IPR003302; Ribosomal_L24_NusG.
 DR Pfam: PF00467; Ribosomal_L24; 1.
 DR ProDom: PD001677; Ribosomal_L24; 1.
 DR ProSITE: PS01108; Ribosomal_L24; FALSE_NEG.
 DR Ribosomal protein; Chloroplast.
 KW Ribosomal protein; Chloroplast.
 SQ SEQUENCE 101 AA; 11572 MW; 306E2E229C636131 CRC64;

 Query Match 2.4%; Score 6; DB 1; Length 101;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 105 SNSLIL 110
 DB 30 SNSLIL 35
 |||||

 RESULT 48
 RFA3_SCHPO STANDARD; PRT; 104 AA.
 AC Q92374;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Replication factor-A protein 3 (Single-stranded DNA-binding protein
 DE p12 subunit).
 GN SSB3 OR SPC23B6.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=972;
 RX MEDLINE=96355433; PubMed=8702843;
 RA Ishai M., Sanchez J.P., Amin A., Murakami Y., Hurwitz J.;
 RT "Purification, gene cloning, and reconstitution of the heterotrimeric
 RT single-stranded DNA-binding protein from Schizosaccharomycetes pombe";
 RL J. Biol. Chem. 271:20868-20878(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Lyne M., Brown S., Quail M., Harris D., Rajandream M.A.,
 RA Barrell B.G.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: BINDS TO SINGLE-STRANDED SEQUENCES.
 CC -!- SUBUNIT: HETEROTRIMER OF 68, 30, AND 12 KDa CHAINS.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U59387; AAC49439.1;
 DR EMBL: AL109608; CAB51564.1;

KW DNA replication; Nuclear protein.
 SQ SEQUENCE 104 AA; 11793 MW; 377BCF913652F1AE CRC64;

 Query Match 2.4%; Score 6; DB 1; Length 104;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 46 TVDNTL 51
 DB 48 TVDNTL 53
 |||||

 RESULT 49
 Y20L_SYNY3 STANDARD; PRT; 109 AA.
 AC P72983;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ycf20-like protein.
 GN SLL1509.
 OS Synechocystis sp. (Strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirose M., Sugiura M., Sugita K., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions";
 RL DNA Res 3:109-136(1996).
 CC -!- SIMILARITY: BELONGS TO THE YCF20 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90902; BAA17002.1;
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 109 AA; 12655 MW; F353C5EDC854EE63 CRC64;

 Query Match 2.4%; Score 6; DB 1; Length 109;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 181 ISLSLL 186
 DB 28 ISLSLL 33
 |||||

 RESULT 50
 HV01_HETER STANDARD; PRT; 116 AA.
 AC P03983;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ig heavy chain V region precursor.
 OS Heterodontus francisci (Horn shark).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
 OC Heterodontidae; Heterodontus.

```

OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A. (CLONE HXIA).
RX MEDLINE=85166276; PubMed=3920659;
RA Litman G.W., Berger L., Murphy K., Litman R., Hinds K.,
RA Erickson B.W.;
RT "Immunoglobulin VH gene structure and diversity in Heterodontus, a
RT phylogenetically primitive shark."
RL Proc. Natl. Acad. Sci. U.S.A. 82:2082-2086(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12195; AAA49326.1; -.
DR PIR; A02086; HVRK.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
KW Immunoglobulin V region; Signal.
FT SIGNAL 1 19
FT CHAIN 20 116 IG HEAVY CHAIN V REGION.
FT DOMAIN 20 49 FRAMEWORK-1.
FT DOMAIN 50 54 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 55 68 FRAMEWORK-2.
FT DOMAIN 69 84 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 85 116 FRAMEWORK-3.
FT DISULFID 41 114 BY SIMILARITY.
FT NON_TER 116 116
SQ SEQUENCE 116 AA; 13114 MW; 3FF573F04E64D4E4 CRC64;

```

```

Query Match      2.4%; Score 6; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 181 ISLSLL 186
   |||||
Db 5 ISLSLL 10

```

Search completed: August 6, 2002, 16:49:56
Job time: 223 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:46:08 ; Search time 26.46 Seconds
(without alignments)
1627.957 Million cell updates/sec

Title: US-10-020-139-2
Perfect score: 249
Sequence: 1 MQLMKLVLLCGVLTGTSSES.....NVIOQVVDNPOHKTQLTLI 249

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database : SPTREMBL19:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rotent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	249	100.0	249	4 Q9BQ00	Q9bq00 homo sapien
2	171	68.7	249	4 Q96DR5	Q96dr5 homo sapien
3	10	4.0	240	6 P79125	P79125 bos taurus
4	10	4.0	243	6 P79124	P79124 bos taurus
5	8	3.2	140	5 Q95SC6	Q95sc6 drosophila
6	8	3.2	353	10 Q9ASR0	Q9asr0 arabidopsis
7	8	3.2	432	3 Q60142	Q60142 schizosacch
8	3.2	525	5 Q9VZD7	Q9vzd7 drosophila	
9	8	3.2	562	4 Q96J40	Q96j40 homo sapien
10	8	3.2	562	11 Q9DOF9	Q9dof9 mus musculu
11	8	3.2	705	16 Q9KQY8	Q9kqy8 vibrio chol
12	8	3.2	720	5 Q62162	Q62162 caenorhabdi
13	8	3.2	817	16 Q9KNF1	Q9knf1 vibrio chol
14	8	3.2	895	10 Q9LX29	Q9lx29 arabidopsis
15	8	3.2	1419	5 Q9V7K0	Q9v7k0 drosophila
16	7	2.8	74	4 Q96A10	Q96a10 homo sapien

ALIGNMENTS

RESULT 1
Q9BQ00
ID Q9BQ00 PRELIMINARY: PRT: 249 AA.
AC Q9BQ00
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BA49G10.1 (SIMILAR TO BOVINE SALIVARY PROTEIN BSP30).
GN BA49G10.1

17	2.8	107	12	Q39489	Q39489 bovine ephe
18	2.8	136	6	Q9MYT3	Q9myt3 oryctolagus
19	2.8	141	17	Q59516	Q59516 pyrococcus
20	2.8	171	4	Q912R7	Q912r7 homo sapien
21	2.8	171	4	Q96RT2	Q96rt2 homo sapien
22	2.8	181	5	Q27202	Q27202 tetrahymena
23	2.8	187	5	Q9VAZ2	Q9vaz2 drosophila
24	2.8	195	9	Q9AYV3	Q9ayv3 bacterioph
25	2.8	199	16	Q92R09	Q92r09 rhizobium m
26	2.8	207	16	P57534	P57534 buchnera ap
27	2.8	208	4	Q9Y5L7	Q9y5l7 homo sapien
28	2.8	212	17	Q973R1	Q973r1 sulfolobus
29	2.8	217	11	Q9EPN2	Q9epn2 mus musculu
30	2.8	222	5	Q95WY6	Q95wy6 ixodes scap
31	2.8	223	11	Q9DD14	Q9dd14 mus musculu
32	2.8	229	4	Q96FT6	Q96ft6 homo sapien
33	2.8	232	2	Q52553	Q52553 amycolatops
34	2.8	235	11	Q63471	Q63471 rattus norv
35	2.8	249	10	Q94FX0	Q94fx0 glycine max
36	2.8	250	10	Q94FX1	Q94fx1 glycine max
37	2.8	255	3	Q9Y7A9	Q9y7a9 metarhizium
38	2.8	257	16	Q97G00	Q97g00 clostridium
39	2.8	262	16	Q927C3	Q927c3 chlamydia p
40	2.8	262	16	Q9J8Y0	Q9jy0 chlamydia p
41	2.8	267	3	Q93187	Q93187 saccharomyc
42	2.8	290	5	Q23394	Q23394 caenorhabdi
43	2.8	295	5	Q9V6U5	Q9v6u5 drosophila
44	2.8	295	16	Q9CNA1	Q9cna1 pasteurella
45	2.8	305	2	Q68121	Q68121 rhodobacter
46	2.8	314	17	Q978D9	Q978d9 thermoplas
47	2.8	322	16	P71065	P71065 bacillus su
48	2.8	369	1	Q9UWQ9	Q9uwq9 thermococcu
49	2.8	373	3	Q9C2D5	Q9c2d5 neurospora
50	2.8	374	16	Q92X04	Q92x04 rhizobium m
51	2.8	381	2	Q53907	Q53907 streptomyce
52	2.8	383	16	Q9W293	Q9w293 thermotoga
53	2.8	389	16	Q9K7L3	Q9k7l3 bacillus ha
54	2.8	390	2	P70928	P70928 bacillus ii
55	2.8	409	16	Q25046	Q25046 helicobacte
56	2.8	409	16	Q92M38	Q92mg8 helicobacte
57	2.8	409	16	Q9PJ52	Q9pj52 campylobact
58	2.8	431	10	Q94E25	Q94e25 oryza sativ
59	2.8	435	17	Q74083	Q74083 pyrococcus
60	2.8	438	10	Q04807	Q04807 perilla fru
61	2.8	442	4	Q9BRK3	Q9brk3 homo sapien
62	2.8	444	10	Q9S229	Q9s229 arabidopsis
63	2.8	448	17	Q59022	Q59022 methanococc
64	2.8	449	10	Q9LJB4	Q9ljb4 arabidopsis
65	2.8	450	4	Q96KC3	Q96kc3 homo sapien
66	2.8	455	3	Q9UTB9	Q9utb9 schizosacch
67	2.8	455	17	Q976E4	Q976e4 sulfolobus
68	2.8	458	16	Q9KXK9	Q9kxk9 vibrio chol
69	2.8	466	3	Q9USN2	Q9usn2 schizosacch
70	2.8	486	10	Q92VH5	Q92vh5 arabidopsis
71	2.8	500	2	Q93JY3	Q93jy3 pseudomonas
72	2.8	502	5	Q9VDM0	Q9vdm0 drosophila
73	2.8	518	2	P71126	P71126 campylobact
74	2.8	533	17	Q973V9	Q973v9 sulfolobus
75	2.8	558	10	Q49624	Q49624 arabidopsis

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL121901; CAC03546.1; -; E64E0794A1B4D87D CRC64;
SQ SEQUENCE 249 AA; 27011 MW; E64E0794A1B4D87D CRC64;

Query Match 100.0%; Score 249; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 2e-246;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLQWLKLVLCGVLGTGTSLLDNLGNDLSNVVVDKLEPVLHGLETVDTNTLKGILEKLV 60
DB 1 MLQWLKLVLCGVLGTGTSLLDNLGNDLSNVVVDKLEPVLHGLETVDTNTLKGILEKLV 60
QY 61 DLGVLQSSAWQAKQAKQAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDGG 120
DB 61 DLGVLQSSAWQAKQAKQAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDGG 120
QY 121 KGLNLSFPVTANVTAGPIIGQIINLKASLDLLTAVTIETDPQTHOPVAVLGECA 180
DB 121 KGLNLSFPVTANVTAGPIIGQIINLKASLDLLTAVTIETDPQTHOPVAVLGECA 180
QY 181 ISLSLLDKHSQIINKFVNSVINTLKVSSLLQKEICPLIRIFHSLLDNNVQQVVDN 240
DB 181 ISLSLLDKHSQIINKFVNSVINTLKVSSLLQKEICPLIRIFHSLLDNNVQQVVDN 240
QY 241 HKTOLOTLI 249
DB 241 HKTOLOTLI 249

RESULT 2
Q96DR5 PRELIMINARY; PRT; 249 AA.
AC Q96DR5;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE PAROTID SECRETORY PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Venkatesh S.G., Gotha C., Gorr S.-U.;
RL "A member of the PSP/plunc family of BPI proteins is expressed in the human parotid gland."
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF432917; AAL28113.1; -;
SQ SEQUENCE 249 AA; 27110 MW; FD54B624A1A4CA7C CRC64;

Query Match 68.7%; Score 171; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.4e-166;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLQWLKLVLCGVLGTGTSLLDNLGNDLSNVVVDKLEPVLHGLETVDTNTLKGILEKLV 60
DB 1 MLQWLKLVLCGVLGTGTSLLDNLGNDLSNVVVDKLEPVLHGLETVDTNTLKGILEKLV 60
QY 61 DLGVLQSSAWQAKQAKQAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDGG 120
DB 61 DLGVLQSSAWQAKQAKQAEKLLNNVSKLLPTNTDIFGLKISNLSLIDVKAEPIDGG 120
QY 121 KGLNLSFPVTANVTAGPIIGQIINLKASLDLLTAVTIETDPQTHOPVAVL 171

```

```

DB 121 KGLNLSFPVTANVTAGPIIGQIINLKASLDLLTAVTIETDPQTHOPVAVL 171

RESULT 3
P79125 PRELIMINARY; PRT; 240 AA.
AC P79125;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE BSP30.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Haigh B.J., Wilkins R.J., Wheeler T.T.;
RL "The cloning and sequencing of two cDNAs coding for alternate forms of BSP30, a bovine member of the parotid secretory protein family."
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U79414; AAB38283.1; -;
SQ SEQUENCE 240 AA; 26513 MW; 850611DE9E43E358 CRC64;

Query Match 4.0%; Score 10; DB 6; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 QLWKLVLVLCG 12
DB 3 QLWKLVLVLCG 12

RESULT 4
P79124 PRELIMINARY; PRT; 243 AA.
AC P79124;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE BSP30.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Haigh B.J., Wilkins R.J., Wheeler T.T.;
RL "The cloning and sequencing of two cDNAs coding for alternate forms of BSP30, a bovine member of the parotid secretory protein family."
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U79413; AAB38282.1; -;
SQ SEQUENCE 243 AA; 26877 MW; 0C2D8DD45660E11C CRC64;

Query Match 4.0%; Score 10; DB 6; Length 243;
Best Local Similarity 100.0%; Pred. No. 0.097;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 QLWKLVLVLCG 12
DB 3 QLWKLVLVLCG 12

RESULT 5
Q95SC6 PRELIMINARY; PRT; 140 AA.
ID Q95SC6

```


AC Q95SC6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE QM03419P.
 GN CG15015.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Scapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Paclet J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY060860; AAL28408.1; -
 SQ SEQUENCE 140 AA; 15502 MW; 6B41E96F24119670 CRC64;

Query Match 3.2%; Score 8; DB 5; Length 140;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 LEKLKVDL 62
 |||||
 Db 28 LEKLKVDL 35

RESULT 6
 Q9ASR0 PRELIMINARY; PRT: 353 AA.
 ID Q9ASR0;
 AC Q9ASR0;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE AT5G62700/MRG21.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Meyers M.C., Shinn P., Banh J.,
 RA Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
 RA Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tortum M., Yamada K., Yu G., Shinozaki K.,
 RA Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF367332; AAK32919.1;
 DR InterPro: IPR002453; Beta-tubulin.
 DR InterPro: IPR000217; Tubulin.
 DR InterPro: IPR003008; Tubulin_ftsz.
 DR Pfam: PF00091; tubulin; 1.
 DR PRINTS: PR01161; TUBULIN.
 DR PROSITE: PS00227; TUBULIN; UNKNOWN_1.
 DR PROSITE: PS00228; TUBULIN_B-AUTOREG; UNKNOWN_1.
 KW GTP-binding.
 SQ SEQUENCE 353 AA; 39251 MW; 2100EE72FCF9F96D CRC64;

Query Match 3.2%; Score 8; DB 10; Length 353;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KLVLGCV 13

Db 336 KLVLGCV 343
 |||||
 RESULT 7
 ID Q60142 PRELIMINARY; PRT: 432 AA.
 AC Q60142;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 50.0 KDA PROTEIN.
 GN SPEC18H10.11c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Lyne M., Rajandream M.A., Barrell B.G., Badcock K., Churcher C.M.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL022304; CAA18408.1; -
 DR InterPro: IPR002885; PPR.
 DR Pfam: PF01535; PPR; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 432 AA; 50048 MW; C395CB5D16C6AE5B CRC64;

Query Match 3.2%; Score 8; DB 3; Length 432;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 VSSLQKE 215
 |||||
 Db 84 VSSLQKE 91

RESULT 8
 Q9VZD7 PRELIMINARY; PRT: 525 AA.
 ID Q9VZD7;
 AC Q9VZD7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG15015 PROTEIN.
 GN CG15015.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M., M.G.,
 RA Palazzolo M., Pittman C.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector R., Turner E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*;"
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003481; AAF47887.1;
 DR FLYBase: FBgn0035533; CG15015.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00346; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 SQ SEQUENCE 525 AA: 58543 MW: 1CE378D9295B64E7 CRC64;

Query Match 3.2% Score 8: DB 5: Length 525;
 Best Local Similarity 100.0%; Pred. No. 22; Indels 0: Gaps 0:
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
 QY 55 LEKLKVDL 62
 DB 349 LEKLKVDL 356
 RESULT 9
 Q96J40 PRELIMINARY; PRT: 562 AA.
 AC Q96J40;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PHOSPHOGLUCUTASE 1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA, AND CHORIOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC001756; AA01756.1;
 SQ SEQUENCE 562 AA: 61370 MW: 8DD05DF242A6F5B4 CRC64;

Query Match 3.2% Score 8: DB 4: Length 562;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
 QY 58 LKVDLGVL 65
 DB 163 LKVDLGVL 170
 RESULT 10
 Q9D0F9 PRELIMINARY; PRT: 562 AA.
 ID Q9D0F9

AC Q9D0F9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 2610020G18RIK PROTEIN.
 GN 2610020G18RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shiuuawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Araiwa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojuna N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 403:685-690(2001).
 DR EMBL: AK011485; BAB27648.1;
 DR HSSP: P00949; 3PMG.
 DR MGD: MGI:1919407; 2610020G18RIK.
 DR InterPro: IPR001485; PGM_PMM.
 DR Pfam: PF00408; PGM_PMM; 1.
 DR PRINTS: PR00509; PGM_PMM.
 DR PROSITE: PS00710; PGM_PMM; 1.
 SQ SEQUENCE 562 AA: 61418 MW: 37854F785A24F71C CRC64;

Query Match 3.2% Score 8: DB 11: Length 562;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
 QY 58 LKVDLGVL 65
 DB 163 LKVDLGVL 170
 RESULT 11
 Q9KQY8 PRELIMINARY; PRT: 706 AA.
 AC Q9KQY8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE METHYL-ACCEPTING CHEMOTAXIS PROTEIN.
 GN VCI859.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,

RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *Cholerae*.";
 RL Nature 406:477-483(2000).
 DR ENBL: AE004261; AAF95007.1; .
 DR HSSP: P02942; 1Q07.
 DR TIGR: VC1859; .
 DR InterPro: IPR004089; Chemotaxis_transducer.
 DR InterPro: IPR003660; HAMP.
 DR Pfam: PF00672; HAMP; 1.
 DR Pfam: PF00015; MCPsignal; 1.
 DR SMART: SM00304; HAMP; 1.
 DR SMART: SM00283; NA; 1.
 KW Complete proteome.
 SQ SEQUENCE 706 AA; 76965 MW; 8C13631C45F769E0 CRC64;

Query Match 3.2%; Score 8; DB 16; Length 706;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 STVSSLQ 213
 |||||
 Db 286 STVSSLQ 293

RESULT 12
 O62162
 ID O62162 PRELIMINARY; PRT; 720 AA.
 AC O62162; Q9U3J4;
 DT 01-AUG-1998 (TremBLrel. 07, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE F14B6.2 PROTEIN.
 GN F14B6.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C.elegans*: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR ENBL: 281502; CAB04106.2; .
 SQ SEQUENCE 720 AA; 81002 MW; FC53702AB0D381B1 CRC64;

Query Match 3.2%; Score 8; DB 5; Length 720;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 EAEKLN 87
 |||||
 Db 286 EAEKLN 293

RESULT 13
 Q9KNF1
 ID Q9KNF1 PRELIMINARY; PRT; 817 AA.
 AC Q9KNF1;
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE MALTODEXTRIN PHOSPHORYLASE.

GN VCA0013.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unwayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RL Nature 406:477-483(2000).
 DR ENBL: AE004345; AAF95927.1; .
 DR HSSP: P00490; 2ECP.
 DR TIGR: VCA0013; .
 DR InterPro: IPR000811; Phosphorylase.
 DR Pfam: PF00343; phosphorylase; 1.
 DR PROSITE: PS00102; PHOSPHORYLASE; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 817 AA; 92565 MW; 253B612226AD722E CRC64;

Query Match 3.2%; Score 8; DB 16; Length 817;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 LKASDLL 153
 |||||
 Db 722 LKASDLL 729

RESULT 14
 Q9LX29
 ID Q9LX29 PRELIMINARY; PRT; 895 AA.
 AC Q9LX29;
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 98.2 KDA PROTEIN.
 RN [25]
 GS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR ENBL: AL356014; CAB91612.1; .
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR004040; STY_pkinase.
 DR InterPro: IPR001368; TNFR_c6.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR SMART: SM00221; TYRKINASE.
 DR SMART: SM00208; TNFR; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108: PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS00050: TNFR_NGFR_2; 1.
 KW ATP-binding: Hypothetical protein; Serine/threonine-protein kinase; Transferrase.
 SQ SEQUENCE 895 AA: 98180 MW: 186BD5B85F0B9DFC CRC64;

Query Match 3.2%; Score 8; DB 10; Length 895;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 CGVLTGTS 18
 |||||
 DB 312 CGVLTGTS 319

RESULT 15
 Q9V7K0 PRELIMINARY: PRT; 1419 AA.

ID Q9V7K0 AC Q9V7K0
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE CG18243 PROTEIN.
 GN CG18243.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophiliidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chert J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hartis N.L., Harvey D., Helman T.J., Hernandez J.K., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Ruskern D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DB EMBL; AF003808; AAF59051.2;

DR HSSP: P18052: LYFO.
 DR Flybase: FBgn0340B5; CG18243.
 DR InterPro: IPR000572; Euk_Oxidored_molyb.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR000387; TYR_phosphatase.
 D4 InterPro: IPR000242; Tyr_prot_phptase.
 DR Pfam: PF00041; fn3; 2.
 DR PRINTS: PRO0700; PRTPHPHTASE.
 DR SMART: SM00060; FN3; 2.
 DR SMART: SM00194; PTPC; 1.
 DR PROSITE: PS00559; MOLYBDOPTERIN_EUK; UNKNOWN_1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 1.
 KW Hydrolase.
 SQ SEQUENCE 1419 AA: 158585 MW: 4719061A0FC5A2C4 CRC64;

Query Match 3.2%; Score 8; DB 5; Length 1419;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 KSTVSSLL 212
 |||||
 DB 644 KSTVSSLL 651

RESULT 16
 Q96A10 PRELIMINARY: PRT; 74 AA.

ID Q96A10 AC Q96A10
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE UNKNOWN (PROTEIN FOR MGC:16737) (PROTEIN FOR MGC:19613).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN, NEUROBLASTOMA, AND MELANOMA;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC011670; AAH11670.1;
 DR EMBL; BC010118; AAH10118.1;
 SQ SEQUENCE 74 AA: 7865 MW: 3E254F3B6D38176B CRC64;

Query Match 2.8%; Score 7; DB 4; Length 74;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 QEAPKLL 85
 |||||
 DB 34 QEAPKLL 40

RESULT 17
 Q39489 PRELIMINARY: PRT; 107 AA.

ID Q39489 AC Q39489
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE BETA PROTEIN.
 OS Bovine ephemeral fever virus (BEFV).
 OC Viruses; SSRNA negative-strand viruses; Mononegavirales;
 OC Rhabdoviridae; Ephemerovirus.
 OX NCBI_TaxID=11303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BEIJING-1;

```

RX MEDLINE=97335257; PubMed=9191923;
RA McWilliam S.M., Kongsuwan K., Cowley J.A., Byrne K.A., Walker P.J.;
RT "Genome organization and transcription strategy in the complex GNS-L
RL intergenic region of bovine ephemeral fever rhabdovirus.";
J. Gen. Virol. 78:1309-1317(1997).
DR ENBL: U72399; AAB63108.1;
SQ SEQUENCE 107 AA; 12219 MW; 4BBZDFBBA0E11709 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 12; Length 107;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 QIINLKA 148
Db 11 QIINLKA 17
|||||

RESULT 18
Q9MYT3 PRELIMINARY; PRT; 136 AA.
AC Q9MYT3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DHAND BASIC HELIX-LOOP-HELIX TRANSCRIPTION FACTOR (FRAGMENT).
GN DHAND.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=HEART;
RA Sayeed R.A., Grace A.A., Vandenberg J.I.;
RT "Immediate-early gene response to acute pressure-overload in the
RT rabbit heart.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DDJ databases.
DR ENBL: AJ291308; CAB94840.1;
DR HSSP; P22415; 1AN4.
DR InterPro: IPR001092; HLH_dim.
DR Pfam: PF00010; HLH; 1.
DR SMART: SM00353; HLH; 1.
FT NON_TER 1
FT NON_TER 136
SQ SEQUENCE 136 AA; 14913 MW; 006512C46A6E0DD6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 6; Length 136;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 LKSTVSS 210
Db 118 LKSTVSS 124
|||||

RESULT 19
O59516 PRELIMINARY; PRT; 141 AA.
AC O59516;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 14.5 KDA PROTEIN PH1841.
GN PH1841.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;

```

```

RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudon Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR ENBL: AP000007; BAA30962.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 141 AA; 14480 MW; 6834B298F6879922 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 17; Length 141;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PTSISLS 184
Db 51 PTSISLS 57
|||||

RESULT 20
Q9Y2R7 PRELIMINARY; PRT; 171 AA.
AC Q9Y2R7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HSPC013
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20499367; PubMed=11042152;
RA Zhang Q.H., Ye M., Wu X.Y., Ren S.X., Zhao M., Zhao C.J., Fu G.,
RA Shen Y., Fan H.Y., Lu G., Zhong M., Xu X.R., Han Z.G., Zhang J.W.,
RA Tao J., Huang Q.H., Zhou J., Hu G.X., Gu J., Chen S.J., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for
RT 300 previously undefined genes expressed in CD34+ hematopoietic
RT stem/progenitor cells.";
RL Genome Res. 10:1546-1560(2000).
DR ENBL: AF077037; AAD27770.1;
DR InterPro: IPR000561; EGF-like.
DR SMART: SM00181; EGF; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 171 AA; 18597 MW; 91A42CD2B2CB0883 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 171;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 TQLQTLI 249
Db 59 TQLQTLI 65
|||||

RESULT 21
Q96RT2 PRELIMINARY; PRT; 171 AA.
AC Q96RT2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P18 PROTEIN.
GN P18.
OS Homo sapiens (Human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RP [1]
 RP SEQUENCE FROM N.A.
 RA Yang Y.C., Chen S.Y., Chang M.S.:
 RT "Cloning and characterization of p18";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275744; AAK69412.1; -;
 SQ SEQUENCE 171 AA; 18567 MW; 81A52CD2B2C008B3 CRC64;

Query Match 2.8%; Score 7; DB 4; Length 171;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 TOLQTLI 249
 Db 59 TOLQTLI 65
 |||||

RESULT 22
 Q27202
 ID Q27202 PRELIMINARY; PRT; 181 AA.
 AC Q27202;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE INITIATION AND CO-STIMULATION FACTOR-A.
 GN ICR-A.
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymenina; Tetrahymena.
 OX NCBI_TaxID=5911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=INRED STRAIN B.
 RA Zhang S., Lockshin C., Orlas E.:
 RT "Nucleotide sequence of the cDNA encoding a mating-related gene from
 Tetrahymena thermophila";
 RL Submitted (JUL-1990) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X53988; CAA37934.1; -;
 SQ SEQUENCE 181 AA; 19813 MW; 97711EC2F6113D5 CRC64;

Query Match 2.8%; Score 7; DB 5; Length 181;
 Best Local Similarity 100.0%; Pred. No. 88;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 ISKLLPT 95
 Db 72 ISKLLPT 78
 |||||

RESULT 23
 Q9VAZ2
 ID Q9VAZ2 PRELIMINARY; PRT; 187 AA.
 AC Q9VAZ2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 GN CG12426 PROTEIN.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 FX MEDLINE=20196006; PubMed=107311132;
 SQ SEQUENCE 187 AA; 22353 MW; 09FF20ECF44B6444 CRC64;

RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,
 RA Brandon R.C., Rogers J.H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow K.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.W., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003762; AAP56754.1; -;
 SQ SEQUENCE 187 AA; 20357 MW; C776041F8CC05B39 CRC64;

Query Match 2.8%; Score 7; DB 5; Length 187;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ILTAVTI 158
 Db 14 ILTAVTI 20
 |||||

RESULT 24
 Q9AYY3
 ID Q9AYY3 PRELIMINARY; PRT; 195 AA.
 AC Q9AYY3;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE 22 KDA PROTEIN.
 OS Bacteriophage MB78 (Salmonella typhimurium bacteriophage MB78).
 OC Viruses.
 OX NCBI_TaxID=52971;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gupta L., Chakravorty M.:
 RT "Cloning and characterization of a gene encoding 22 kDa functional
 protein of bacteriophage MB78";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF349435; AAK30159.1; -;
 SQ SEQUENCE 195 AA; 22353 MW; 09FF20ECF44B6444 CRC64;

Query Match 2.8%; Score 7; DB 9; Length 195;

Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 ASLDLLT 154
|||||||
Db 38 ASLDLLT 44

RESULT 25

Q92R09 ID Q92R09 PRELIMINARY; PRT: 199 AA.
AC Q92R09
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE COLICIN V PRODUCTION HOMOLOG TRANSMEMBRANE PROTEIN.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis K.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goureau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hymen R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Rampeger U., Surzycki R., Thebault P., Vandenbol M.,
RA Vorheeler F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.,
RT "The composite genome of the legume symbiont Sinorhizobium meliloti";
RL Science 293:668-672(2001).
DR EMBL; AL591786; CAC45704.1;
KW Complete proteome.
SQ SEQUENCE 199 AA: 21156 MW: FE2438A45ADDIBEF CRC64;

Query Match 2.8%; Score 7; DB 16; Length 199;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LLDNLGN 27
|||||||
Db 144 LLDNLGN 150

RESULT 26

P57534 ID P57534 PRELIMINARY; PRT: 207 AA.
AC P57534
DT 01-WAR-2001 (TREMBLrel. 16, Created)
DT 01-WAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RIBOFLAVIN REDUCTASE.
GN RIBD2 OR BU452.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbolic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.,
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS".
RL Nature 407:81-86(2000).
DR EMBL; AF001119; BAB13159.1;
InterPro; IPR002734; RibD.C.

Pfam: PF01872; RibD.C; 1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 207 AA: 23942 MW: DAC973284089D76A CRC64;

Query Match 2.8%; Score 7; DB 16; Length 207;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LKISNSL 108
|||||||
Db 178 LKISNSL 184

RESULT 27

Q9Y517 ID Q9Y517 PRELIMINARY; PRT: 208 AA.
AC Q9Y517
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE APOPTOSIS RELATED PROTEIN APR-3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhu F., Yan W., Chai Y.B., Shao C., Peng W.D., Yang A.G., Wang C.J.,
RA Zhao Z.L.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144055; AAD31317.2;
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00001; EGF-like; 1.
SQ SEQUENCE 208 AA: 22687 MW: CD4C5723C62CAF6F CRC64;

Query Match 2.8%; Score 7; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 TQLQTLI 249
|||||||
Db 83 TQLQTLI 89

RESULT 28

Q975R1 ID Q975R1 PRELIMINARY; PRT: 212 AA.
AC Q975R1
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE 3-OCTAPHENYL-4-HYDROXYBENZOATE CARBOXY-LYASE.
GN ST0360.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagil M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
Crenarchaeon, Sulfolobus tokodaii strain 7".
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000982; BAB65339.1;
KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 212 AA; 23346 MW; EFAD7B9E11606323 CRC64;

Query Match 2.8%; Score 7; DB 17; Length 212;
Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 102 LKISNSL 108
| | | | |
DB 188 LKISNSL 194

RESULT 29

ID Q9EPN2 PRELIMINARY; PKT; 217 AA.
AC Q9EPN2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE DHAND PROTEIN.
GN DHAND.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kemp P.R., Chen Q., Metcalfe J.C.;
RT "Counter-regulation of Hand and Twist gene family expression is
RT correlated with the differentiated phenotype of vascular smooth muscle
RT cells."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ131846; CAC20671.1;
DR HSSP: P36956; IAM9.
DR InterPro: IPR001092; HLH_dim.
DR Pfam: PF00010; HLH; 1.
DR SMART: SM00353; HLH; 1.
SQ SEQUENCE 217 AA; 23735 MW; A6CA5555FF7173F1F CRC64;

Query Match 2.8%; Score 7; DB 11; Length 217;
Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 204 LKSTVSS 210
| | | | |
DB 189 LKSTVSS 195

RESULT 30

ID Q95WY6 PRELIMINARY; PKT; 222 AA.
AC Q95WY6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 25 KDA SALIVARY GLAND PROTEIN A.
GN SAI25A.
OS Ixodes scapularis (black-legged tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Ixodes.
OX NCBI_TaxID=6945;
RN [1]
RP SEQUENCE FROM N.A.
RA Das S., Banerjee G., DePonte K., Marcantonio N., Kantor F.S.,
RA Fikrig E.;
RT "Salp25D, an Ixodes scapularis antioxidant, is one of 14
RT immunodominant antigens in engorged tick salivary glands."
RL J. Infect. Dis. 184:0-0(2001).
DR EMBL: AF209922; AAK97825.1;
SQ SEQUENCE 222 AA; 25368 MW; 4D1A6062763FEEC0 CRC64;

Query Match 2.8%; Score 7; DB 5; Length 222;
Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 8 VLLCGVL 14
| | | | |
DB 9 VLLCGVL 15

RESULT 31

ID Q9DU14 PRELIMINARY; PKT; 223 AA.
AC Q9DU14;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 0610007C21RIK PROTEIN.
GN 0610007C21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-C57BL/6J; TISSUE-KIDNEY;
RX MEDLINE-21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schrim L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK002276; BAB21981.1;
DR MGI: 1918918; 0610007C21RIK.
DR InterPro: IPR000561; EGF-like.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00001; EGF-like; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN.1.
SQ SEQUENCE 223 AA; 23867 MW; A85E1A03C70D2C16 CRC64;

Query Match 2.8%; Score 7; DB 11; Length 223;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 TQLQTLI 249
| | | | |
DB 111 TQLQTLI 117

RESULT 32

ID Q96FF6 PRELIMINARY; PKT; 229 AA.
AC Q96FF6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:13322).
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN, AND GLIOBLASTOMA;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC011006; AAH11006.1; -
 SQ SEQUENCE 229 AA; 24688 MW; F2C63F934A47ED33 CRC64;

Query Match 2.8%; Score 7; DB 4; Length 229;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 TQLOTLL 249
 |||||
 DB 117 TQLOTLL 123

RESULT 33
 O52553
 ID O52553 PRELIMINARY; PRT; 232 AA.
 AC O52553;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE PHOSPHATASE.
 GN RIFM.
 OS Amycolatopsis mediterranei (Nocardia mediterranei).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
 OX NCBI_TaxID=33910;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=98174059; PubMed=9512878;
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Yu T.W.,
 RA Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,
 RA Floss H.G.;
 RA "Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
 RT the molecular analysis of the rif biosynthetic gene cluster of
 RT Amycolatopsis mediterranei S699.";
 RL Chem. Biol. 5:69-79(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=98165773; PubMed=9497318;
 RA Kim C.G., Yu T.W., Fryhle C.R., Handa S., Floss H.G.;
 RT "3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
 RT formation of the precursor of mC7N units in rifamycin and related
 RT antibiotics.";
 RL J. Biol. Chem. 273:6030-6040(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=21201076; PubMed=11278540;
 RA Yu T.-W., Mueller R., Mueller M., Zhang X., Draeger G., Kim C.-G.,
 RA Leistner E., Floss H.G.;
 RT "Mutational analysis and reconstituted expression of the biosynthetic
 RT genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the
 RT starter unit of rifamycin biosynthesis in Amycolatopsis mediterranei
 RT S699.";
 RL J. Biol. Chem. 276:12546-12555(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,
 RA Floss H.G.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF040570; AAC01721.1; -
 DR InterPro: IPR001454; Hydrolase.

DR Pfam: PF00702; Hydrolase; 1.
 SQ SEQUENCE 232 AA; 24854 MW; 7A977F7A550DBF15 CRC64;

Query Match 2.8%; Score 7; DB 2; Length 232;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 SLJNLG 26
 |||||
 DB 129 SLJNLG 135

RESULT 34
 O63471
 ID O63471 PRELIMINARY; PRT; 235 AA.
 AC O63471;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE NEONATAL SUBMANDIBULAR GLAND PROTEIN PRECURSOR.
 GN PSP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
 RX MEDLINE=92129360; PubMed=1370829;
 RA Mireis L., Ball W.D.;
 RA "Neonatal rat submandibular gland protein SMG-A and parotid secretory
 RT protein are alternatively regulated members of a salivary protein
 RT multigene family.";
 RL J. Biol. Chem. 267:2679-2687(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
 RX MEDLINE=98129760;
 RA Mireis L., Miranda A.J., Ball W.D.;
 RT "Characterization of the rat salivary-gland B1-immunoreactive
 RT proteins.";
 RL Biochem. J. 330:437-444(1998).
 DR EMBL: M83209; AAC06334.1; -
 KW Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 235 NEONATAL SUBMANDIBULAR GLAND PROTEIN.
 SQ SEQUENCE 235 AA; 24529 MW; 0B36EC779025986E CRC64;

Query Match 2.8%; Score 7; DB 11; Length 235;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GTSESLL 22
 |||||
 DB 16 GTSESLL 22

RESULT 35
 Q94FX0
 ID Q94FX0 PRELIMINARY; PRT; 249 AA.
 AC Q94FX0;
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE HEME OXYGENASE 3 (FRAGMENT).
 GN HO3.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurasids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=21295560; PubMed=11402195;
RA Davis S.J., Bhoo S.H., Durski A.M., Walker J.M., Vierstra R.D.;
RT "The heme-oxygenase family required for phytochrome chromophore
RT biosynthesis is necessary for proper photomorphogenesis in higher
RT plants.";
RL Plant Physiol. 126:656-669(2001).
DR EMBL; AF320025; AAK63009.1; -.
FT NON_TER 1
SQ SEQUENCE 249 AA; 28319 MW; 203BE0A22C455E22 CRC64;

Query Match 2.8%; Score 7; DB 10; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 AEKLLN 87
DB 185 AEKLLN 191

RESULT 36
Q94FX1 PRELIMINARY; PRT; 250 AA.
AC Q94FX1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEME OXYGENASE 1 (FRAGMENT).
OS HOI.
GN HOI.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21295560; PubMed=11402195;
RA Davis S.J., Bhoo S.H., Durski A.M., Walker J.M., Vierstra R.D.;
RT "The heme-oxygenase family required for phytochrome chromophore
RT biosynthesis is necessary for proper photomorphogenesis in higher
RT plants.";
RL Plant Physiol. 126:656-669(2001).
DR EMBL; AF320024; AAK63008.1; -.
FT NON_TER 1
SQ SEQUENCE 250 AA; 28544 MW; E0241C35ED1F5F1F CRC64;

Query Match 2.8%; Score 7; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 AEKLLN 87
DB 186 AEKLLN 192

RESULT 37
Q9Y7A9 PRELIMINARY; PRT; 255 AA.
AC Q9Y7A9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPSIN-RELATED PROTEASE.
GN TRY2.
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
OX NCBI_TaxID=5530;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-ME1;
RA Screen S.E., St Leger R.J.;
RT "Isolation of multiple protease genes from the entomopathogenic fungus
RT Metarhizium anisopliae.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF130865; AAD29675.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.103; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRY-SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 255 AA; 26289 MW; 88DU979ED300E457 CRC64;

Query Match 2.8%; Score 7; DB 3; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 ANVTVAG 137
DB 146 ANVTVAG 152

RESULT 38
Q97G00 PRELIMINARY; PRT; 257 AA.
ID Q97G00;
AC Q97G00;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PREDICTED S-ADENOSYLMETHIONINE-DEPENDENT METHYLTRANSFERASE.
GN CAC2574.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Moeelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007755; AAK80523.1; -.
DR InterPro; IPR000780; ChR_methtransf.
DR InterPro; IPR001601; Meth-transf.
DR InterPro; IPR000051; SAM_bind.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 257 AA; 29989 MW; 4E09E5DE2D8CD09B CRC64;

Query Match 2.8%; Score 7; DB 16; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 EAEKLLN 86
DB 158 EAEKLLN 164

RESULT 39

```

```

Q927C3
ID Q927C3 PRELIMINARY: PRT: 262 AA.
AC Q927C3
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CTS98 HYPOTHETICAL PROTEIN.
GN CPN0783.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
DR ENBL: AB001659; AAD18921.1;
KW Complete proteome.
SQ SEQUENCE 262 AA; 29010 MW; AD80B2FBF22DB8FE CRC64;

Query Match 2.8%; Score 7; DB 16; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 KTQLQTL 248
Db 132 KTQLQTL 138
|||||

RESULT 40
Q927C3
ID Q927C3 PRELIMINARY: PRT: 262 AA.
AC Q927C3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CTS98 HYPOTHETICAL PROTEIN.
GN CPJ0783 OR CP1089;
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Head T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR ENBL: AB002265; AAF38860.1;
DR EMBL: AP002547; BAA98991.1;
DR TIGR: CP1089;
SQ SEQUENCE 262 AA; 29019 MW; AD80B2FBFBC31FE CRC64;

Query Match 2.8%; Score 7; DB 16; Length 262;

Q927C3
ID Q927C3 PRELIMINARY: PRT: 262 AA.
AC Q927C3
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CHROMOSOME VI LAMBDA CLONE.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Churchill C.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrell H., Rajandream M.A.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
DR ENBL: Z46255; CAA86348.1;
DR InterPro: IPR002114; PTS_HPT_ser.
DR PROSITE: PS00589; PTS_HPT_SER; UNKNOWN_1.
SQ SEQUENCE 267 AA; 28771 MW; ACA79F341A4319FD CRC64;

Query Match 2.8%; Score 7; DB 3; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 STVSSLL 212
Db 203 STVSSLL 209
|||||

RESULT 42
Q23394
ID Q23394 PRELIMINARY: PRT: 290 AA.
AC Q23394
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ZK1086.3 PROTEIN.
GN ZK1086.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR ENBL: Z70269; CAA94222.1;
DR InterPro: IPR002485; DUF13.
DR Pfam: PF01482; DUF13.1;
DR ProDom: PD002726; DUF13.1;

```

SQ SEQUENCE 250 AA; 33362 MW; 80856303FBCF4D6F CRC64;

Query Match 2.8%; Score 7; DB 5; Length 290;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 ILEKLKV 60

DB 142 ILEKLKV 148

RESULT 43

Q9V6U5

ID Q9V6U5 PRELIMINARY; PRT: 295 AA.

AC Q9V6U5

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE CG18645 PROTEIN

GN CG5543 OR CG18645

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20195006; PubMed=107311132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yangell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer J.R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.N., Miklos G.L.G.,

RA April J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.A., Bouck J., Brockstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeqam C.,

RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Koudry C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,

RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinett K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*."

RL Science 287:2185-2195(2000).

DR EMBL: AE003817; AAF58327.1;

DR HSSP: P14604; 2DUB.

DR FlyBase: FB00033879; CG5543.

DR InterPro: IPR001753; EnOyl_CoA_hydrtse.

DR Pfam: PF00378; ECH; 1.

DR PROSITE: PS00166; ENOYL_COA_HYDRATASE; 1.

SQ SEQUENCE 295 AA; 31562 MW; 82C853BDBEF705B9 CRC64;

Query Match 2.8%; Score 7; DB 5; Length 295;

Best Local Similarity 100.0%; Pred. No. 1.4e+02; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 AOEAEKL 84

DB 205 AOEAEKL 211

RESULT 44

Q9CNA1

ID Q9CNA1 PRELIMINARY; PRT: 295 AA.

AC Q9CNA1

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE ISPA.

GN ISPA OR PM0533.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Pasteurella.

OX NCBI_TaxID=747;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PM70;

RX MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RA "Complete genomic sequence of *Pasteurella multocida* PM70."

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

DR EMBL: AE006088; AAK02617.1;

DR InterPro: IPR000092; Polyprenyl_synth.

DR Pfam: PF00348; Polyprenyl_synth. 1.

DR PROSITE: PS00444; POLYPRENYL_SYNTHET_2; 1.

KW Complete proteome.

SQ SEQUENCE 295 AA; 32151 MW; F90SC60F2A504644 CRC64;

Query Match 2.8%; Score 7; DB 16; Length 295;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 AKQKAQE 80

DB 258 AKQKAQE 264

RESULT 45

O68121

ID O68121 PRELIMINARY; PRT: 305 AA.

AC O68121

DT 01-AUG-1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE RIBOSE TRANSPORT ATP-BINDING PROTEIN.

OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).

OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;

OC Rhodobacter.

OX NCBI_TaxID=1061;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SB1003;

RX MEDLINE=9740404; PubMed=9256491;

RA Vleck C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;

RA "Sequence of a 189-kb segment of the chromosome of *Rhodobacter*

RA capsulatus SB1003."

RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).

DR EMBL: AF010496; AAC16211.1;

DR InterPro: IPR003593; AAA.

DR InterPro: IPR003439; ABC_transportr.

```

DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00005; ABC_tran; 1.
DR SMART: SM00382; AAA; 1.
KW ATP-binding
SQ SEQUENCE 305 AA: 32914 MW: 55AD58A707E922F0 CRC64:

Query Match      2.8%; Score 7; DB 2: Length 305;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 LILDVKA 114
    |||||
Db 193 LILDVKA 199

RESULT 46
Q978D9 PRELIMINARY; PRT; 314 AA.
AC Q978D9
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE FORMATE HYDROGENLYASE SUBUNIT 5.
GN TVG1529492
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaaceae;
OC Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570456; PubMed=11121031;
RA Kawashima T., Anano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshima T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL: AP000996; BAB60520.1;
KW Lyase, Complete proteome.
SQ SEQUENCE 314 AA: 35061 MW: 2DA07090F36892ED CRC64:

Query Match      2.8%; Score 7; DB 17: Length 314;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 GILEKLK 59
    |||||
Db 25 GILEKLK 31

RESULT 47
P71065 PRELIMINARY; PRT; 322 AA.
ID P71065
AC P71065; O08183;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 37.3 KDA PROTEIN.
GN YVFF
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168TRP;
RA Fabret C., Quentin Y., Chapel N., Guiseppe A., Halech J., Denizot F.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

InterPro: IPR001687; ATP_GTP_A.
Pfam: PF00005; ABC_tran; 1.
SMART: SM00382; AAA; 1.
ATP-binding
SEQUENCE 305 AA: 32914 MW: 55AD58A707E922F0 CRC64:

Query Match      2.8%; Score 7; DB 2: Length 305;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 LILDVKA 114
    |||||
Db 193 LILDVKA 199

RESULT 46
Q978D9 PRELIMINARY; PRT; 314 AA.
AC Q978D9
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE FORMATE HYDROGENLYASE SUBUNIT 5.
GN TVG1529492
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaaceae;
OC Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570456; PubMed=11121031;
RA Kawashima T., Anano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshima T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL: AP000996; BAB60520.1;
KW Lyase, Complete proteome.
SQ SEQUENCE 314 AA: 35061 MW: 2DA07090F36892ED CRC64:

Query Match      2.8%; Score 7; DB 17: Length 314;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 GILEKLK 59
    |||||
Db 25 GILEKLK 31

RESULT 47
P71065 PRELIMINARY; PRT; 322 AA.
ID P71065
AC P71065; O08183;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 37.3 KDA PROTEIN.
GN YVFF
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168TRP;
RA Fabret C., Quentin Y., Chapel N., Guiseppe A., Halech J., Denizot F.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

Denizot F.C.;
Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98044033; PubMed=9384377;
RX Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
RA Avevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschii C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton J.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Enrich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: 271928; CAA96483.1;
DR EMBL: 294043; CAB07999.1;
DR EMBL: 299121; CAB15427.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 322 AA: 37274 MW: CE0B1FF05A498E73 CRC64:

Query Match      2.8%; Score 7; DB 16: Length 322;
Best Local Similarity 100.0%; Pred. No. 1.5e-02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 209 SSLQKE 215
    |||||
Db 275 SSLQKE 281

RESULT 48
Q9UNQ9 PRELIMINARY; PRT; 369 AA.
ID Q9UNQ9
AC Q9UNQ9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NIFE HYDROGENASE BETA SUBUNIT.
GN HYDB.
OS Thermococcus litoralis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI_TaxID=2265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 4573;

```

RX MEDLINE=20050631; PubMed-10583413;
 RA Rakheyl G., Zhou Z.Z., Adams M.W.W., Kovacs K.L.;
 RT "Biochemical and molecular characterization of the [NifH] hydrogenase
 from the hyperthermophilic archaeon, Thermococcus litoralis";
 RL Eur. J. Biochem. 266:1158-1165(1999).
 DR EMBL: AF039208; AAB94933.1; -.
 DR InterPro: IPR001450; 4Fe4S_ferredoxin.
 DR Pfam: PF00037; fer4; 1.
 DR PROSITE: PS00198; 4Fe4S_FKRED0XIN; 2.
 SQ SEQUENCE 369 AA: 43293 MW: 42332FFB7D749973 CRC64;

Query Match 2.8%; Score 7; DB 1; Length 369;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 DIFGLKI 104
 |||||
 DB 99 DIFGLKI 105

RESULT 49

Q9C2G5 PRELIMINARY; PRT; 373 AA.
 AC Q9C2G5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 41.0 KDA PROTEIN.
 GN 93G11.280.
 OS Neurospora crassa.
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 CC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL513443; CAC28680.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 373 AA: 40973 MW: 7A9704F940E22DB9 CRC64;

Query Match 2.8%; Score 7; DB 3; Length 473;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 STVSLL 212
 |||||
 DB 323 STVSLL 329

RESULT 50

Q92X04 PRELIMINARY; PRT; 374 AA.
 AC Q92X04;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN SMB20158.
 GN SMB20158.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 CC Plasmid pSymB (megaplasmid 2).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=1021;
 RX MEDLINE=21396508; PubMed-11481431;
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Couzy J.,
 RA Golding B., Puehler A.;
 RT "The complete sequence of the 1.683-kb pSynB megaplasmid from the N2-
 fixing endosymbiont Sinorhizobium meliloti";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 DR EMBL: AL603642; CAC48558.1; -.
 KW Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 374 AA: 40663 MW: 3B5CA5574EBB3797 CRC64;

Query Match 2.8%; Score 7; DB 16; Length 374;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 LPTNTDI 99
 |||||
 DB 323 LPTNTDI 329

Search completed: August 6, 2002, 16:49:40
 Job time: 212 sec

KW Human: cancer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnerary; antilucer; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

OS Homo sapiens.

XX WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-457603/49.

XX N-PSDB; AAH99686.

PT Isolated human polynucleotides encoding polypeptides, useful for the
 treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

PS Claim 20; Page 260; 1217pp; English.

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
 CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.

SQ Sequence 260 AA;

Query Match 53.8%; Score 134; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 2.6e-115;
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLQLKLVLLCGVLTGTSSELDNLGNDLSNVVDKLEPVLHGLETVNTLKGILEKLV 60
 DB 12 mlqlwlvllcgvltgtsesllldnldnsnvvdleplvhegletdntlkglekikv 71

QY 61 DLGVLOKSSAWOLAKOAKQAEKLLANNVISKLLPTNTDIFGLKISNLSILDVKAEPIDG 120
 DB 72 dlglvqkssawqlakqakeakllnnviskllptndifgklsnlsildvkaepidg 131
 QY 121 KGLNLSFPVTANVT 134
 DB 132 kglnlsfpvtanvt 145

RESULT 7

ID ABB41435 standard; Peptide; 50 AA.

XX ABB41435;

DT 04-FEB-2002 (first entry)

DE Peptide #8941 encoded by human foetal liver single exon probe.

KW Human: foetal liver; gene expression: single exon nucleic acid probe.

OS Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -

XX Claim 27; SEQ ID NO 34070; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at fip.wipo.int/pub/published_pct_sequences.

XX Sequence 50 AA;

Query Match 20.1%; Score 50; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.6e-38;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 PIIGQIINLKASLDLITAVTIETDPOTHQPVAVLGECDPTISISLID 187
 DB 1 pligqinlkasldlitavtiectdpqthqpvavigeacdptsislild 50

RESULT 8

AAM62308

ID AAM62308 standard; Protein; 50 AA.

```

XX AC AAM62308;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 34413.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX FA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Kank DR;
XX XX WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains.
XX PS Example 4; SEQ ID NO: 34413; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention.
XX SQ Sequence 50 AA:

Query Match 20.1%; Score 50; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.6e-38;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 PIIGQINLKASLDLLTAVTETDPQTHQPVAVLGECASDPTISLSLLD 187
Db 1 pligqinlkasldlltavtietdpqthqpvavlgecasdptsislid 50

RESULT 9
AAM75111
ID AAM75111 standard; Protein; 50 AA.
XX AC AAM75111;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 35417.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX XX

Query Match 20.1%; Score 50; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.6e-38;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 PIIGQINLKASLDLLTAVTETDPQTHQPVAVLGECASDPTISLSLLD 187
Db 1 pligqinlkasldlltavtietdpqthqpvavlgecasdptsislid 50

RESULT 9
AAM75111
ID AAM75111 standard; Protein; 50 AA.
XX AC AAM75111;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 35417.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX XX

Query Match 20.1%; Score 50; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.6e-38;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 PIIGQINLKASLDLLTAVTETDPQTHQPVAVLGECASDPTISLSLLD 187
Db 1 pligqinlkasldlltavtietdpqthqpvavlgecasdptsislid 50

RESULT 10
AAM35227
ID AAM35227 standard; Protein; 50 AA.
XX AC AAM35227;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #9264 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX XX

```

```

PN WO200157276-A2.
PD 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US00668.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Kank DR;
XX XX WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO: 35417; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX SQ Sequence 50 AA:

Query Match 20.1%; Score 50; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.6e-38;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 PIIGQINLKASLDLLTAVTETDPQTHQPVAVLGECASDPTISLSLLD 187
Db 1 pligqinlkasldlltavtietdpqthqpvavlgecasdptsislid 50

RESULT 10
AAM35227
ID AAM35227 standard; Protein; 50 AA.
XX AC AAM35227;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #9264 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX XX

```

PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-48897/53.
 DR Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta.
 PT Claim 27; SEQ ID No 35496; 654pp; English.
 PS The present invention relates to single exon nucleic acid probes (SENP;
 XX see AAI31315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX Sequence 50 AA:
 SQ

Query Match 20.1%; Score 50; DB 22; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.6e-38;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 PIIGQIINKASIDLLTAVTETDPTQHPVAVLGECDPTSLSLD 187
 DB 1 PIIGQIINKASIDLLTAVTETDPTQHPVAVLGECDPTSLSLD 50

RESULT 11
 ABC20804
 ID ABC20804 standard; Protein; 316 AA.
 XX AC ABG20804;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #20795.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS84991.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID No 51163; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes, (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABC00010-ABC30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 316 AA:
 SQ

Query Match 3.2%; Score 8; DB 22; Length 316;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LKVDLGLVL 65
 DB 109 LKVDLGLVL 116

RESULT 12
 AAB54100
 ID AAB54100 standard; Protein; 511 AA.
 XX AC AAB54100;
 XX DT 09-MAR-2001 (first entry)
 XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:552.
 XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.
 XX OS Homo sapiens.
 XX PN WC2000055320-A1.
 XX PD 21-SEP-2000.
 XX PF 08-MAR-2000; 2000WO-US05989.
 XX PR 12-MAR-1999; 99US-0124270.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM;
 XX DR WPI; 2000-579444/54.
 XX DR N-PSDB; AAC98865.
 XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition.
 XX Claim 11; Page 990-992; 1379pp; English.
 XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,

neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive, gynaecological, cardiac and antiinflammatory activities, and can be used in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer. Agonists and antagonists to the antigens can be screened for. The pancreatic cancer antigen polynucleotides can be used to design nucleic acid hybridisation probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both in vivo and in vitro diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent sequences used in the exemplification of the present invention.

XX SQ Sequence 511 AA;

Query Match 3.2%; Score 8; DB 21; Length 511;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LKVDLGLV 65
| | | | | | | |
DB 155 lkvdllglv 162

RESULT 13

ABE70056
ID ABE70056 standard; Protein: 525 AA.

XX AC ABE70056;
XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 36960.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX XX (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI: 2001-656860/75.

XX DR N-PSDB; ABL14159.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from *Drosophila* and for elucidating cell signalling and cell-cell interactions.

XX PS Disclosure; SEQ ID NO 36960; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABE57737-ABE72072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 525 AA;

Query Match 3.2%; Score 8; DB 22; Length 525;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 LEKLKVDL 62
| | | | | | | |
DB 349 lekikvdl 356

RESULT 14

AAU58382
ID AAU58382 standard; Protein: 601 AA.

XX AC AAU58382;

XX DT 27-FEB-2002 (first entry)

XX DE Propionibacterium acnes immunogenic protein #19278.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis; uveitis; endophthalmitis; bone; joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US12855.

XX PR 21-APR-2000; 2000US-199047P.

XX PR 02-JUN-2000; 2000US-208841P.

XX PR 07-JUL-2000; 2000US-216747P.

XX XX (CORI-) CORIXA CORP.

XX PA Skeiky YAW, Persing DH, Mitcham Jr, Wang SS, Bhatia A;

XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX DR WPI: 2001-616774/71.

XX DR N-PSDB; AAS59590.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris.

XX PS Example 1; SEQ ID NO 19577; 1069pp; English.

XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies

CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 601 AA;

Query Match 3.2%; Score 8; DB 22; Length 601;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 PIIGQIIN 145
 Db 263 pligqiin 270
 |||||

RESULT 15
 ABB71134
 ID ABB71134 standard; Protein; 1419 AA.

XX AC ABB71134;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 40194.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEXE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL15237.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX PS Disclosure: SEQ ID NO 40194; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA
 CC sequences (AB101840-ABL16175) and the encoded proteins
 CC (AB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1419 AA;

Query Match 3.2%; Score 8; DB 22; Length 1419;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 KSTVSSL 212
 Db 644 kstvsall 651
 |||||

RESULT 16

AA41392

ID AAY41392 standard; Protein; 43 AA.

XX AC AAY41392;

XX DT 02-DEC-1999 (first entry)

XX DE Human secreted protein encoded by gene 85 clone HSDGN55.

XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX PN WO9947540-A1.

XX PD 23-SEP-1999.

XX PF 18-MAR-1999; 99WO-US05804.

XX PR 19-MAR-1998; 98US-0078563.

XX PR 19-MAR-1998; 98US-0078566.

XX PR 19-MAR-1998; 98US-0078573.

XX PR 19-MAR-1998; 98US-0078574.

XX PR 19-MAR-1998; 98US-0078576.

XX PR 19-MAR-1998; 98US-0078577.

XX PR 19-MAR-1998; 98US-0078578.

XX PR 19-MAR-1998; 98US-0078579.

XX PR 19-MAR-1998; 98US-0078581.

XX PR 01-APR-1998; 98US-0080312.

XX PR 01-APR-1998; 98US-0080313.

XX PR 01-APR-1998; 98US-0080314.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;

XX PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW;

XX PI Olsen HS, Shi Y, Moore PA;

XX DR WPI; 1999-562050/47.

XX DR N-PSDB; AA224895.

XX PT New isolated human genes, useful for diagnosis and treatment of e.g.

XX PT cancers, neurological disorders, immune diseases, inflammation or blood

XX PT disorders.

XX PS Claim 11: Page 408; 484pp; English.

XX CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AA224802) for increasing the stability of the fused
 CC protein as compared to the human protein only.

XX CC The invention relates to 95 novel genes and their fragments (nucleic
 CC acid sequences: AA224811-224907; amino acid sequences AAY41308-Y41404)

XX CC which are useful for preventing, treating or ameliorating medical

XX CC conditions e.g. by protein or gene therapy. Also, pathological


```

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX N-PSDB; AAK58271.
XX
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Claim 11; SEQ ID NO 13083; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 46 AA:
XX
XX Query Match 2.8%; Score 7; DB 22; Length 46;
XX Best Local Similarity 100.0%; Pred. No. 63;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 80 EAEKLLN 86
XX | | | | |
XX Db 24 eaeklln 30

```

```

RESULT 18
AAK85206
ID AAK85206 standard; Protein; 59 AA.
XX
AC AAK85206;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen SEQ ID NO:12799.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 11-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.

```

```

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-483426/52.
XX N-PSDB: AAK57987.
DR
DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Claim 11; SEQ ID NO 12799; 3071pp + Sequence Listing; English.
PS
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 59 AA;
SQ
Query Match 2.8%; Score 7; DB 22; Length 59;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 54 ILEKLKV 60
Db 42 ILEKLKV 48
RESULT 19
AAAG57800
ID AAG57800 standard; Protein; 61 AA.
XX
AC AAG57800;
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 74530.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
XX
XX

```

[illegible]

PR 23-SEP-1999; 990S-0155486.
 PR 24-SEP-1999; 990S-0155659.
 PR 28-SEP-1999; 990S-0156458.
 PR 29-SEP-1999; 990S-0156596.
 PR 04-OCT-1999; 990S-0157117.
 PR 05-OCT-1999; 990S-0157753.
 PR 06-OCT-1999; 990S-0157865.
 PR 07-OCT-1999; 990S-0158029.
 PR 08-OCT-1999; 990S-0158232.
 PR 12-OCT-1999; 990S-0158369.
 PR 13-OCT-1999; 990S-0159293.
 PR 13-OCT-1999; 990S-0159294.
 PR 13-OCT-1999; 990S-0159295.
 PR 14-OCT-1999; 990S-0159329.
 PR 14-OCT-1999; 990S-0159330.
 PR 14-OCT-1999; 990S-0159637.
 PR 14-OCT-1999; 990S-0159638.
 PR 18-OCT-1999; 990S-0159584.
 PR 21-OCT-1999; 990S-0160741.
 PR 21-OCT-1999; 990S-0160767.
 PR 21-OCT-1999; 990S-0160768.
 PR 21-OCT-1999; 990S-0160770.
 PR 21-OCT-1999; 990S-0160814.
 PR 21-OCT-1999; 990S-0160815.
 PR 22-OCT-1999; 990S-0160980.
 PR 22-OCT-1999; 990S-0160981.
 PR 22-OCT-1999; 990S-0160989.
 PR 23-OCT-1999; 990S-0161404.
 PR 25-OCT-1999; 990S-0161405.
 PR 25-OCT-1999; 990S-0161406.
 PR 26-OCT-1999; 990S-0161359.
 PR 26-OCT-1999; 990S-0161360.
 PR 26-OCT-1999; 990S-0161361.
 PR 28-OCT-1999; 990S-0161320.
 PR 28-OCT-1999; 990S-0161932.
 PR 28-OCT-1999; 990S-0161933.
 PR 29-OCT-1999; 990S-0162142.

Query Match 2.8%; Score 7; DB 21; Length 61;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 204 LKSTVSS 210

Db 46 lkstvs 52

RESULT 20

ABBI6925

ID ABB16925 standard; Protein: 65 AA.

AC ABB16925;

XX 23-JAN-2002 (first entry)

XX Human nervous system related polypeptide SEQ ID NO 5582.

DE Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;

KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;

KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;

KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;

KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;

KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;

KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

OS Homo sapiens.

XX WC200159063-A2.

PN 16-AUG-2001.

PD 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

PF 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216847.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234223.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.

PR	02-OCT-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	02-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241826;
PR	20-OCT-2000;	2000US-0242221;
PR	01-NOV-2000;	2000US-0246417;
PR	08-NOV-2000;	2000US-0246574;
PR	08-NOV-2000;	2000US-0246575;
PR	08-NOV-2000;	2000US-0246476;
PR	08-NOV-2000;	2000US-0246477;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0246523;
PR	08-NOV-2000;	2000US-0246524;
PR	08-NOV-2000;	2000US-0246525;
PR	08-NOV-2000;	2000US-0246526;
PR	08-NOV-2000;	2000US-0246527;
PR	08-NOV-2000;	2000US-0246528;
PR	08-NOV-2000;	2000US-0246532;
PR	08-NOV-2000;	2000US-0246600;
PR	08-NOV-2000;	2000US-0246611;
PR	08-NOV-2000;	2000US-0246613;
PR	17-NOV-2000;	2000US-0246613;
PR	17-NOV-2000;	2000US-0249207;
PR	17-NOV-2000;	2000US-0249208;
PR	17-NOV-2000;	2000US-0249209;
PR	17-NOV-2000;	2000US-0249210;
PR	17-NOV-2000;	2000US-0249211;
PR	17-NOV-2000;	2000US-0249212;
PR	17-NOV-2000;	2000US-0249213;
PR	17-NOV-2000;	2000US-0249215;
PR	17-NOV-2000;	2000US-0249216;
PR	17-NOV-2000;	2000US-0249217;
PR	17-NOV-2000;	2000US-0249237;
PR	17-NOV-2000;	2000US-0249239;
PR	01-DEC-2000;	2000US-0249300;
PR	01-DEC-2000;	2000US-0250391;
PR	05-DEC-2000;	2000US-0251160;
PR	05-DEC-2000;	2000US-0251030;
PR	05-DEC-2000;	2000US-0251988;
PR	06-DEC-2000;	2000US-0256719;
PR	06-DEC-2000;	2000US-0251479;
PR	08-DEC-2000;	2000US-0251856;
PR	08-DEC-2000;	2000US-0251868;
PR	08-DEC-2000;	2000US-0251869;
PR	08-DEC-2000;	2000US-0251989;
PR	08-DEC-2000;	2000US-0251990;
PR	11-DEC-2000;	2000US-0259497;
PR	05-JAN-2001;	2001US-0259587;

useful for preventing, diagnosing and/or treating nervous system
cancers and metastases -

Claim 11: SEQ ID NO 5582; 170lpp + Sequence Listing; English.

The invention relates to novel genes (ABAI1004-ABA21534) and proteins
(ABBI4678-ABBI18001) useful for preventing, treating or ameliorating
medical conditions e.g. by protein or gene therapy. The genes are
isolated from a range of human tissues disclosed in the specification.
The nucleic acids, proteins, antibodies and (ant)agonists are useful
in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
and ovarian cancer and other cancers of the adrenal gland, bone, bone
marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
(b) immune disorders e.g. Addison's disease, allergies, autoimmune;
haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
disease, multiple sclerosis, rheumatoid arthritis and ulcerative
colitis; (c) cardiovascular disorders such as myocardial ischaemias;
(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
and parasitic infections.

Note: The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from Wipo at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 65 AA:

PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 12-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237017.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 03-DEC-2000; 2000US-0251030.
 PR 03-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM:

WPI; 2001-451929/48.
 N-PSDB; AAS40209.

Isolated polypeptide for treating, preventing and/or prognosing disorders related to the reproductive system including prostate cancer and also for testing and detection e.g. diagnosis -

Claim 11; SEQ ID No 361; 546pp; English.

The invention relates to novel isolated human prostate cancer antigen polynucleotides (I) and polypeptides (II). (I) and (II) are useful for preventing, treating or ameliorating a medical condition when administered. (I), (II) and the antibody to (II) are useful for treating, preventing and/or prognosing disorders related to the reproductive system including prostate cancers; urinary disorders e.g. chronic nephritis; and blood-related disorders e.g. thrombosis. (II) can be used for testing and detection e.g. as a chromosomal marker and in forensics. (I) and the anti-(II) antibody can be used in testing and detection in immunoassays. AAU22702-AAU22913 represent the human prostate cancer antigen amino acid sequences, and related amino acid sequences of the

CC invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic
CC directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 80 AA;

Query Match 2.8%; Score 7; DB 22; Length 80;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 DIFGLKI 104
|||||
Db 38 difglki 44

RESULT 22
AAM96148
ID AAM96148 standard; Protein; 80 AA.

XX AC AAM96148;
XX DT 21-NOV-2001 (first entry)

XX DE Human reproductive system related antigen SEQ ID NO: 4806.

XX KW Human reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01339.

XX PR 31-JAN-2000; 2000US-0173065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226581.

XX PR 22-AUG-2000; 2000US-0226588.

XX PR 23-AUG-2000; 2000US-0227182.

XX PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-02595678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 DR N-PSDB; AAL02118.
 XX

PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Claim 11; SEQ ID NO 4806; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a protein of the invention.
 XX

SQ Sequence 80 AA;

Query Match 2.8%; Score 7; DR 22; Length 80;
 Best Local Similarity 100.0%; Pred. No. 1.e-02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 DIFGLKI 104
 DB 38 dlfglki 44

RESULT 23
 ABB38333
 ID ABB38333 standard; Peptide; 86 AA.
 XX
 AC ABB38333;
 XX

DT 04-FEB-2002 (first entry)

DE Peptide #5839 encoded by human foetal liver single exon probe.
 XX Human; foetal liver; gene expression; single exon nucleic acid probe.
 KW Homo sapiens.
 OS
 XX WO200157277-A2.
 PN
 XX

PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Kank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -

XX Claim 27; SEQ ID NO 30968; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 86 AA;

Query Match 2.8%; Score 7; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.e-02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 LTGTSES 20
 DB 79 ltgtses 85

RESULT 24

ARM58947
 ID ARM58947 standard; Protein; 86 AA.
 XX

AC ARM58947;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 31052.
 DE Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX

OS Homo sapiens.

XX WO200157275-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483446/52.
 DR Single exon nucleic acid probes for analyzing gene expression in human
 XX brains -
 PT Example 4: SEQ ID NO: 31052; 650pp + Sequence Listing; English.
 XX
 PS The present invention provides a number of single exon nucleic acid
 XX probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 XX the probes of the invention.
 XX
 SQ Sequence 86 AA;

Query Match 2.8%; Score 7; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LTGTSES 20
 |||||
 Db 79 lgtsses 85

RESULT 25
 AAM31774
 ID AAM31774 standard; Protein; 86 AA.
 XX
 AC AAM31774;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #5811 encoded by probe for measuring placental gene expression.
 XX
 DE Probe: microarray; human; placenta; antenatal diagnosis;
 XX genetic disorder.
 KW
 KW Homo sapiens.
 XX
 OS WO200157272-A2.
 XX
 PN 09-AUG-2001.
 XX
 PD 30-JAN-2001; 2001WO-US00663.
 XX
 PF 04-FEB-2000; 2000US-0180312.
 XX
 PR 26-MAY-2000; 2000US-0207456.
 XX
 PR 30-JUN-2000; 2000US-0608408.
 XX
 PR 03-AUG-2000; 2000US-0632366.
 XX
 PR 21-SEP-2000; 2000US-0234687.
 XX
 PR 27-SEP-2000; 2000US-0236359.
 XX
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PT Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID No 32043; 654pp; English.
 XX

CC The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 86 AA;

Query Match 2.8%; Score 7; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LTGTSES 20
 |||||
 Db 79 lgtsses 85

RESULT 26
 AAG08597
 ID AAG08597 standard; Protein; 92 AA.
 XX
 AC AAG08597;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 6197.
 XX
 DE Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 XX
 PR 05-MAR-1999; 99US-0123180.
 XX
 PR 09-MAR-1999; 99US-0123548.
 XX
 PR 23-MAR-1999; 99US-0125788.
 XX
 PR 25-MAR-1999; 99US-0126264.
 XX
 PR 29-MAR-1999; 99US-0126785.
 XX
 PR 01-APR-1999; 99US-0127462.
 XX
 PR 06-APR-1999; 99US-0128234.
 XX
 PR 08-APR-1999; 99US-0128714.
 XX
 PR 16-APR-1999; 99US-0129845.
 XX
 PR 19-APR-1999; 99US-0130077.
 XX
 PR 21-APR-1999; 99US-0130449.
 XX
 PR 23-APR-1999; 99US-0130510.
 XX
 PR 28-APR-1999; 99US-0130891.
 XX
 PR 28-APR-1999; 99US-0131449.
 XX
 PR 30-APR-1999; 99US-0132048.
 XX
 PR 30-APR-1999; 99US-0132407.
 XX
 PR 04-MAY-1999; 99US-0132484.
 XX
 PR 05-MAY-1999; 99US-0132485.
 XX
 PR 06-MAY-1999; 99US-0132486.
 XX
 PR 06-MAY-1999; 99US-0132487.
 XX
 PR 07-MAY-1999; 99US-0132863.
 XX
 PR 11-MAY-1999; 99US-0134256.
 XX
 PR 14-MAY-1999; 99US-0134218.
 XX
 PR 14-MAY-1999; 99US-0134219.
 XX
 PR 14-MAY-1999; 99US-0134221.
 XX
 PR 14-MAY-1999; 99US-0134221.
 XX
 PR 18-MAY-1999; 99US-0134370.
 XX
 PR 19-MAY-1999; 99US-0134768.
 XX
 PR 19-MAY-1999; 99US-0134941.
 XX
 PR 20-MAY-1999; 99US-0135124.
 XX
 PR 21-MAY-1999; 99US-0135353.
 XX
 PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141387.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161559.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

PR 28 OCT-1999; 99US-0161993.
PR 29 OCT-1999; 99US-0162142.
Query Match 2.8%; Score 7; DB 21; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 26 GNDLSNV 32
Db 31 gndlsnv 37
RESULT 27
AAG08596
ID AAG08596 standard; Protein: 95 AA.
XX AC AAG08596;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEO ID NO: 6196.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135144.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.

Query Match 2.8%; Score 7; DB 21; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels

Qy	26	GNDSNV 32	
Db	34	gnalsnv 40	
RESULT 28			
	AAV73496		
ID	AAV73496 standard: Protein; 103 AA.		
XX	AC	AAV73496;	
XX	DT	29-FEB-2000 (first entry)	
XX	XX	Human secreted protein clone ysl0_1 protein sequence SEQ ID NO:214.	
XX	XX	Human; secreted protein; immunostimulatory; haemostatic; cytokine; proliferative; differentiative; chemotactic; chemokinetic; vaccine; thrombolytic; antinflammatory; cytostatic; immunosuppressive;	
KW	KW	gene therapy.	
XX	XX	Homo sapiens.	
XX	OS	XX	
XX	PN	WO9586642-A2.	
XX	PD	18-NOV-1999.	
XX	XX	XX	
XX	PF	14-MAY-1999; 99WO-US10843.	
XX	XX	14-MAY-1998; 98US-0085472.	
PR	17-AUG-1998; 98US-0096824.		
PR	11-SEP-1998; 98US-0099843.		
PR	11-SEP-1998; 98US-0099950.		
PR	15-SEP-1998; 98US-0100424.		
PR	29-SEP-1998; 98US-0102329.		
PR	09-OCT-1998; 98US-0103615.		
PR	11-DEC-1998; 98US-0111799.		
PR	14-DEC-1998; 98US-0112159.		
PR	31-DEC-1998; 98US-0114415.		
PR	10-FEB-1999; 99US-0248059.		
PR	06-APR-1999; 99US-0287150.		
XX	13-MAY-1999; 99US-0311021.		
XX	(GEMY) GENETICS INST INC.		
PA	Wong GG, Clark HF, Fecthel K, Agostino MJ;		
PI	WPI: 2000-053095/04.		
DR	N-PSDR: AA252581.		
DR	Novel polynucleotides and proteins having biological activities which make them suitable for treating, preventing or ameliorating medical conditions in humans or animals -		
XX	Claim 223; Page 719; 730pp: English.		
XX	The present invention describes human secreted proteins encoded by polynucleotides obtained from adult testes, foetal brain, adult brain, brain (foetal and adult), foetal kidney, adult spleen, and adult thymus cDNA libraries. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals.		
CC	Suggested activities include nutritional activity, cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotides are also stated to be useful for gene therapy. Therapeutic compositions are also presently valuable for veterinary applications. AA252475 to AA252581 encode human secreted proteins, and AAV73390 to AAV73500 represent human secreted proteins, given in the present invention.		

```
SQ Sequence 103 AA:
Query Match 2.8%; Score 7; DB 21; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 KASLDLL 153
Db 4 kasidll 10

RESULT 29
AAG08595
ID AAG08595 standard; Protein; 110 AA.
XX
AC AAG08595;
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6195.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
KW
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132046.
PR 03-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134216.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137526.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
```

```
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159395.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Qy 26 GNDLSNV 32
Db 49 gndlsnv 55

RESULT 30
ABG04836
ID ABG04836 standard; Protein: 113 AA.
XX ABG04836;
AC ABG04836;
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4827.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDE: AAS69023.
XX
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.
Claim 20: SEQ ID NO 35195; 103pp: English.
XX
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Query Match 2.8%; Score 7; DB 22; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 QAEKLL 85
```

Db 43 qeak11 49
|||||

RESULT 31
ABG12485
ID ABG12485 standard; Protein: 130 AA.

XX AC ABG12485;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #12476.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB: AAS76672.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity

XX PS Claim 20; SEQ ID No 42844; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 130 AA;

Query Match 2.8%; Score 7; DB 22; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 LDLTAV 156
|||||

Db 94 ldltav 100

RESULT 32

ABG19852
ID ABG19852 standard; Protein: 130 AA.

XX AC ABG19852;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #19843.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB: AAS84039.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity

XX PS Claim 20; SEQ ID No 50211; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 130 AA;

Query Match 2.8%; Score 7; DB 22; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 LDLTAV 156
|||||

RESULT 33
 AAC00157
 ID AAC00157 standard; Protein; 147 AA.
 XX AC AAG00157;
 XX AC AAG00157;
 XX DT 06-OCT-2000 (first entry)
 XX DE Human secreted protein, SEQ ID NO: 4238.
 XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX OS Homo sapiens.
 XX PN EP1033401-A2.
 XX PD 06-SEP-2000.
 XX PF 21-FEB-2000; 2000EP-0200610.
 XX PR 26-FEB-1999; 99US-0122487.
 XX PA (GIST) GENSET.
 XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.
 XX N-PSDB: AAC00163.
 XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 13; SEQ ID 4238; 71pp + CD-ROM; English.
 XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX SQ Sequence 147 AA;
 Query Match 2.8%; Score 7; DB 21; Length 147;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 243 TQLQTLI 249
 Db 117 tqqltli 123
 RESULT 34
 AAU19625
 ID AAU19625 standard; Protein; 162 AA.
 XX AC AAU19625;
 XX AC AAU19625;
 XX DT 04-DEC-2001 (first entry)
 XX DE Human diagnostic and therapeutic polypeptide (DITHP) #211.

XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KW respiratory disorder.
 XX OS Homo sapiens.
 XX PN WO200162927-A2.
 XX PD 30-AUG-2001.
 XX PF 21-FEB-2001; 2001WO-US06059.
 XX PR 24-FEB-2000; 2000US-0184693.
 PR 24-FEB-2000; 2000US-0184697.
 PR 24-FEB-2000; 2000US-0184698.
 PR 24-FEB-2000; 2000US-0184768.
 PR 24-FEB-2000; 2000US-0184769.
 PR 24-FEB-2000; 2000US-0184770.
 PR 24-FEB-2000; 2000US-0184771.
 PR 24-FEB-2000; 2000US-0184772.
 PR 24-FEB-2000; 2000US-0184773.
 PR 24-FEB-2000; 2000US-0184774.
 PR 24-FEB-2000; 2000US-0184776.
 PR 24-FEB-2000; 2000US-0184777.
 PR 24-FEB-2000; 2000US-0184797.
 PR 24-FEB-2000; 2000US-0184813.
 PR 24-FEB-2000; 2000US-0184837.
 PR 24-FEB-2000; 2000US-0184841.
 PR 24-FEB-2000; 2000US-0185213.
 PR 24-FEB-2000; 2000US-0185216.
 PR 12-MAY-2000; 2000US-0203785.
 PR 15-MAY-2000; 2000US-0204226.
 PR 16-MAY-2000; 2000US-0204525.
 PR 16-MAY-2000; 2000US-0204821.
 PR 16-MAY-2000; 2000US-0204908.
 PR 16-MAY-2000; 2000US-0205232.
 PR 17-MAY-2000; 2000US-0204815.
 PR 17-MAY-2000; 2000US-0204863.
 PR 17-MAY-2000; 2000US-0205221.
 PR 17-MAY-2000; 2000US-0205285.
 PR 17-MAY-2000; 2000US-0205286.
 PR 17-MAY-2000; 2000US-0205287.
 PR 17-MAY-2000; 2000US-0205323.
 PR 17-MAY-2000; 2000US-0205324.
 XX (INCY-) INCYTE GENOMICS INC.
 XX PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'Sa SA, Ansley S, Dahl CR, Dam TC, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
 PI Wright RJ, Yap PB, Yu JY, Bradley DL, Bratcher SR, Chen W,
 PI Cohen RJ, Hodgson DM, Lincoln SE, Jackson S;
 XX WPI: 2001-502867/55.
 DR N-PSDB: AAS31196.
 XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
 PT
 PT
 XX Claim 27; Page 522; 522pp; English.
 XX The invention relates to polynucleotides (I) encoding diagnostic and
 CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
 CC and proteins involved in growth and development and receptors. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate DITHP expression. For example, (i) and
 CC (ii) may be used to treat disorders associated with decreased polypeptide
 CC expression by rectifying mutations or deletions in a patient's genome,
 CC that affect the activity of the DITHPs, by expressing inactive proteins

CC or supplementing the patient's own production of them. (I) and (II)
 CC may be used to treat diseases, for example, cell proliferative disorder,
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
 CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
 CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. (II) may also be used as antigens in the production of
 CC antibodies against DITHPs and in assays to identify modulators of DITHP
 CC expression and activity. The anti-DITHP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DITHP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DITHPs in samples (e.g., by enzyme linked immunosorbent
 CC assay (ELISA)). AAU19415-AAU19525 represent human diagnostic and
 CC therapeutic (DITHP) polypeptides of the invention.
 XX
 SQ Sequence 162 AA;

Query Match 2.8%; Score 7; DB 22; Length 162;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 243 TQLQTLI 249
 Db 114 tqqltli 120
 |||||

RESULT 35

AAU19282
 ID AAU19282 standard; Protein; 168 AA.

XX
 AC AAU19282;

XX
 DT 04-DEC-2001 (first entry)

XX
 DE Human G protein-coupled receptor nGPCR-2459.

XX Human; G protein-coupled receptor; nGPCR-x; antiviral; analgesic;
 KW cytosolic; cardiac; antidiabetic; anorectic; hypotensive; hypertensive;
 KW antiparkinsonian; neurotrophic; neuroprotective; antidepressant;
 KW viral infection; HIV-1; human immunodeficiency virus; HIV-2; pain;
 KW cancer; metabolic disease; cardiovascular disease; type 2 diabetes;
 KW obesity; anorexia; hypotension; hypertension; myocardial infarction;
 KW atherosclerosis; parkinson's disease; psychosis; neurological disorder;
 KW schizophrenia; migraine; major depression; anxiety; mental disorder;
 KW manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.

XX Homo sapiens.

XX WO200166750-A2.

XX PD 13-SEP-2001.

XX PF 08-MAR-2001; 2001WO-US07322.

XX XX 08-MAR-2000; 2000US-0187581.

XX PR 08-MAR-2000; 2000US-0187582.

XX PR 08-MAR-2000; 2000US-0187714.

XX PR 08-MAR-2000; 2000US-0187715.

XX PR 08-MAR-2000; 2000US-0187825.

XX PR 08-MAR-2000; 2000US-0187828.

XX PR 08-MAR-2000; 2000US-0187829.

XX PR 08-MAR-2000; 2000US-0187830.

XX PR 08-MAR-2000; 2000US-0187833.

XX PR 08-MAR-2000; 2000US-0187874.

XX PR 08-MAR-2000; 2000US-0187930.

XX PR 08-MAR-2000; 2000US-0188049.

XX PR 08-MAR-2000; 2000US-0189294.

XX PR 08-MAR-2000; 2000US-0187929.

XX PR 08-MAR-2000; 2000US-0187928.

XX (PHAA) PHARMACIA & UPJOHN CO.

XX PI Vogel G, Wood LS;

XX DR WPI: 2001-536778/59.

XX DR N-PSDB: AAS30851.

XX Isolated nucleic acid molecules encoding G protein-coupled receptors
 PT termed nGPCR-x, useful in the treatment and diagnosis of viral
 PT infections, cancers and mental disorders (e.g. Parkinson's disease and
 PT schizophrenia) -

XX PS Claim 31; Page 327; 336pp; English.

XX The invention relates to novel isolated nucleic acid molecules encoding
 CC G protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides,
 CC polypeptides, and modulators may be used in the treatment of diseases and
 CC conditions such as infections, such as viral infections caused by HIV-1
 CC (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and
 CC cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,
 CC anorexia, hypotension, hypertension, myocardial infarction,

CC atherosclerosis), Parkinson's disease, and psychotic and
 CC neurological disorders, including schizophrenia, migraine, major
 CC depression, anxiety, mental disorder, manic depression, and
 CC dyskinesias, such as Huntington's disease or Tourette's Syndrome
 CC and many other diseases and syndromes listed in the specification.
 CC nGPCR-x polynucleotides and polypeptides, as well as nGPCR-x
 CC modulators, may also be used in diagnostic assays for such diseases or
 CC conditions. The present sequence represents a G protein-coupled
 CC receptor of the invention.

XX SQ Sequence 168 AA;

Query Match 2.8%; Score 7; DB 22; Length 168;

Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 STVSSL 212
 Db 19 stvssl 25
 |||||

RESULT 36

ABB67511
 ID ABB67511 standard; Protein; 187 AA.

XX AC ABB67511;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 29325.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX XX 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI: 2001-656860/75.

DR N-PSDB; ABL11614.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions.
 PS Disclosure; SEQ ID NO 29325; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABU01840-ABU16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 187 AA;
 PS
 Query Match 2.8%; Score 7; DB 22; Length 187;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 152 LLTAVTI 158
 DB 14 lltavti 20
 RESULT 37
 AAY35414
 ID AAY35414 standard; Protein; 194 AA.
 XX
 AC AAY35414;
 XX
 DT 13-SEP-1999 (first entry)
 DE Amino acid sequence of a *Chlamydia pneumoniae* protein.
 XX
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX
 OS *Chlamydia pneumoniae*.
 XX
 PN WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 XX 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI; 1999-357842/30.
 XX
 XX Genome sequence of *Chlamydia pneumoniae*
 PT
 PS Page 1198; Disclosure; 1912pp; English.
 XX
 XX AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of *Chlamydia pneumoniae*.
 CC *C. pneumoniae* causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the *C. pneumoniae* genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing *C. pneumoniae*
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of *C. pneumoniae*.
 XX
 XX Sequence 194 AA;
 PS
 Query Match 2.8%; Score 7; DB 20; Length 194;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 242 KTLQLOTL 248
 DB 71 ktqlqtl 77
 RESULT 38
 AAU25572
 ID AAU25572 standard; Protein; 203 AA.
 XX
 AC AAU25572;
 XX
 DT 18-DEC-2001 (first entry)
 DE Human G Protein-Coupled Receptor (GPCR) polypeptide #19.
 XX
 KW Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
 KW attention deficit disorder; anxiety; depression; bipolar disorder;
 KW neurological disorder; Huntington's disease; dementia; obesity; anorexia;
 KW metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
 KW type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
 KW cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
 KW viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;
 KW antidepressant; anorectic; gene therapy.
 XX
 OS *Homo sapiens*.
 XX
 PN WO200162797-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 23-FEB-2001; 2001WO-US05676.
 XX
 PR 23-FEB-2000; 2000US-0184247.
 PR 23-FEB-2000; 2000US-0184303.
 PR 23-FEB-2000; 2000US-0184304.
 PR 23-FEB-2000; 2000US-0184305.
 PR 23-FEB-2000; 2000US-0184397.
 PR 02-MAR-2000; 2000US-0186457.
 PR 03-MAR-2000; 2000US-0186810.
 PR 09-MAR-2000; 2000US-0188064.
 PR 13-MAR-2000; 2000US-0188880.
 PR 03-APR-2000; 2000US-0194344.
 PR 23-JUN-2000; 2000US-0213861.
 PR 11-JUL-2000; 2000US-0217359.
 PR 11-JUL-2000; 2000US-0217370.
 PR 14-JUL-2000; 2000US-0218337.
 PR 20-JUL-2000; 2000US-0218492.
 XX
 XX (PAAA) PHARMACIA & UPJOHN CO.
 XX
 PI Vogeli G, Wood LS, Parodi LA, Lind P;
 XX
 DR WPI; 2001-570628/64.
 XX
 DR N-PSDB; AAS42824.
 XX
 XX New isolated nucleic acid encoding a new G-protein coupled receptor
 PT polypeptide for detecting receptor modulators that can treat mental
 PT disorders, such as schizophrenia, anxiety, depression, or obesity -
 XX
 PS Claim 35; Page 77; 279pp; English.

XX Sequences AAU25554-AAU25616 represent human G-protein coupled receptor
 CC (GPCR) polypeptides of the invention. The proteins and their associated
 CC DNA sequences can be used to identify compounds which bind to GPCR
 CC polypeptides and in screening for compounds that modulate GPCR activity.
 CC By screening a human subject for the presence of mutations in GPCR DNA, a
 CC GPCR-related disorder or a genetic predisposition can be diagnosed. The
 CC sequences can also be used for treatment and prevention of mental
 CC disorders such as schizophrenia, attention deficit disorder, anxiety,
 CC depression, dementia and bipolar disorder. neurological disorders such as
 CC Huntington's disease, Parkinson's disease and Tourette's syndrome,
 CC metabolic disorders such as obesity, anorexia and type 2 diabetes,
 CC cardiovascular disorders such as thrombosis, myocardial infarction,
 CC cardiomyopathy and atherosclerosis, viral infections caused by HIV and
 CC cancers.

XX Sequence 203 AA;

Query Match 2.8%; Score 7; DB 22; Length 203;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LKISNSL 108
 DB 193 LKISNSL 199
 |||||

RESULT 39
 AAU29361
 ID AAU29361 standard; Protein: 211 AA.
 XX AC AAU29361;
 XX DT 18-DEC-2001 (first entry)
 XX DE Novel mar regulated protein (NIMR) #33.
 XX KW mar regulated polypeptide; NIMR; microbial infection; antibacterial.
 XX OS Escherichia coli.
 XX PN WO200170776-A2.
 XX PD 27-SEP-2001.
 XX PF 08-MAR-2001; 2001WO-US07478.
 XX PR 10-MAR-2000; 2000US-188362P.
 XX PA (TUFT) TUFTS COLLEGE.
 XX PI Levy SB, Barbosa TM, Alekshun MN;
 XX WPI: 2001-602769/68.
 XX DR N-PSDB; AAS46261.
 XX PT Identifying compounds that modulate a newly identified mar regulated
 PT polypeptide activity, useful as antimicrobial compounds, involves
 PT contacting the polypeptide with a test compound.
 XX PS Disclosure; Page 382-383; 526pp; English.

The invention relates to a method of identifying compounds that modulate
 CC a newly identified mar regulated (NIMR) polypeptide activity. The method
 CC comprises contacting an NIMR polypeptide with a test compound under
 CC interaction conditions, determining the ability of the compound to
 CC modulate the activity or expression of the polypeptide, and selecting the
 CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
 CC of microbial infections, and in screening for modulators of NIMR
 CC expression and activity. These modulators can be used to reduce the
 CC infectivity of a microbe on a surface, and the virulence of a microbe in
 CC a subject suffering from an infection. AAU29329-AAU29379 represent

CC Escherichia coli NIMK amino acid sequences of the invention.
 XX Sequence 211 AA;
 SQ

Query Match 2.8%; Score 7; DB 22; Length 211;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 GILEKIK 59
 DB 74 gilekik 80
 |||||

RESULT 40
 AAB99925
 ID AAB99925 standard; Protein: 217 AA.
 XX AC AAB99925;
 XX DT 26-SEP-2001 (first entry)
 XX DE Rat dHAND protein sequence SEQ ID NO:21.
 XX KW Differentiation; heart muscle cell; cytokine; transcription factor;
 KW proliferation; surface antigen; heart disease; cardiomyocyte;
 KW bone marrow; umbilical blood cell; heart muscle degeneration;
 KW myocardial infarction.
 XX OS Rattus norvegicus.
 XX PN WO200148150-A1.
 XX PD 05-JUL-2001.
 XX PF 02-NOV-2000; 2000WO-JP07741.
 XX PR 28-DEC-1999; 99JP-0372826.
 XX PR 28-FEB-2000; 2000WO-JP01148.
 XX PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX PI Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;
 XX Yamaoka Y;
 XX WPI: 2001-425655/45.
 XX DR N-PSDB; AAH44361.
 XX PT Cells capable of differentiating into cardiomyocytes and originating in
 PT bone marrow or umbilical blood cells for study of cardiomyocyte
 PT differentiation and treatment of heart disease
 XX PS Claim 46; Page 102-121; 187pp; Japanese.

The present invention describes cells originating in bone marrow or
 CC umbilical blood cells which are capable of differentiating into
 CC cardiomyocytes. Also described are: (1) cardiomyocytes produced by the
 CC differentiation of the cells; (2) a method for carrying out the
 CC differentiation into cardiomyocytes, regulated by a promotional and/or
 CC inhibitory factor; (3) a method for the differentiation of the cells
 CC into cell types other than cardiomyocytes; (4) drug compositions
 CC promoting the formation of heart muscle and regeneration of heart tissue
 CC which contain the cells; (5) a method for the production of antibodies
 CC which recognise the cells, especially antibodies which recognise a
 CC surface antigen on the cells; (6) a method for screening factors which
 CC promote the proliferation of the cells; (7) a method for immortalising
 CC the cells by expressing telomerase in them; (8) drug compositions for
 CC the treatment of heart disease which contain the immortalised cells; and
 CC (9) cell-free supernatant from the culture of the cells and its use in
 CC promoting their differentiation into cardiomyocytes. The cells are used
 CC in the treatment of diseases involving heart muscle degeneration, such
 CC as myocardial infarction and in the study of cardiomyocyte
 CC differentiation. AAH44351 to AAH44409 and AAB99915 to AAB99935 represent

CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 217 AA;

Query Match 2.8%; Score 7; DB 22; Length 217;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 LKSTVSS 210

DB 189 lkstvs 195

RESULT 41

AAG64324

ID AAG64324 standard; Protein; 217 AA.

AC AAG64324;

DT 24-SEP-2001 (first entry)

DE Rat dHAND protein.

KW Angiogenesis; cardiant; cell differentiating agent; bone marrow;

KW heart muscle cell; heart disease; rat; dHAND.

OS Rattus norvegicus.

PN WO200148149-A1.

XX 05-JUL-2001.

PF 28-FEB-2000; 2000WO-JP01148.

PR 28-DEC-1999; 99JP-0372826.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PI Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K;

DR WPI; 2001-418252/44.

DR N-PSDB; AAH49596.

XX New adult bone marrow-originated cells capable of differentiating into

PT heart muscle cells, applicable as remedies for various heart diseases

PT particularly with damaged heart muscle accompanying degeneration

PS Claim 28; Pages 105-106; 158pp; Japanese.

XX The present invention relates to cells isolated from bone marrow, which

CC are capable of at least differentiating into heart muscle cells. The

CC cells are applicable as remedies for various heart diseases particularly

CC with damaged heart muscle accompanying degeneration. The present sequence

CC was used to illustrate the present invention.

XX
SQ Sequence 217 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 LKSTVSS 210

DB 189 lkstvs 195

RESULT 42

AAG64854

ID AAG64854 standard; Protein; 217 AA.

XX
AC AAG64854;

XX

DT

XX

DE

XX

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

XX

PR

XX

PR

XX

PA

XX

XX

PI

XX

PI

XX

XX

DR

XX

DR

XX

XX

PT

XX

PT

XX

XX

XX

PS

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

CC

XX

AC

XX

XX

DT

XX

DE

XX

XX

KW

XX

KW

XX

XX

OS

XX

XX

PN

XX

XX

PD

XX

XX

21-SEP-2001 (first entry)

Heart muscle cell differentiation related protein SEQ ID NO: 21.

Heart muscle cell; human; cell differentiation; heart disease.

Homo sapiens.

WO200148151-A1.

05-JUL-2001.

27-DEC-2000; 2000WO-JP09323.

28-DEC-1999; 99JP-0372826.

28-FEB-2000; 2000WO-JP01148.

02-NOV-2000; 2000WO-JP07741.

(KYOW) KYOWA HAKKO KOGYO KK.

Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;

Yamada Y;

WPI; 2001-425656/45.

N-PSDB; AAH48230.

Cells capable of differentiating into cardiomyocytes and originating in

bone marrow or umbilical blood cells for study of cardiomyocyte

differentiation and treatment of heart disease

Claim 36; Page 126-127; 183pp; Japanese.

The present invention provides cells originating in the human bone marrow

or umbilical blood cells which are capable of differentiating into

cardiomyocytes. These cells are useful in the treatment of diseases

involving heart muscle degeneration, such as myocardial infarction, and

the study of cardiomyocyte differentiation. The present sequence is

a protein described in the exemplification of the invention.

Sequence 217 AA;

Query Match 2.8%; Score 7; DB 22; Length 217;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 LKSTVSS 210

DB 189 lkstvs 195

RESULT 43

AAY97770

ID AAY97770 standard; Protein; 222 AA.

XX

XX

XX

06-AUG-2001 (first entry)

I. scapularis Salp25A protein sequence.

Salp; tick; antibody; tick immunity; infection; tick-borne disease;

coagulation factor Xa inhibitor; inflammatory response.

Ixodes scapularis.

WO200140469-A2.

07-JUN-2001.

01-DEC-2000; 2000WO-US32765.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

PR 03-DEC-1999; 99US-0169048.
PR 16-OCT-2000; 2000US-0240716.
XX (UYVA ) UNIV YALE.
XX Kantor FS, Fikrig E, Das S;
XX WPI; 2001-367810/38.
DR N-PSDB; AAA91501.
XX Novel ixodes scapularis polypeptides for conferring tick immunity and
XX for preventing the transmission of tick-borne pathogens -
XX Claim 20; Fig 16; 137pp; English.
XX This sequence is an ixodes scapularis polypeptide of the invention.
XX The proteins of the invention are 15 tick Salp proteins. The proteins,
XX antibodies against them or pharmaceutical composition comprising a fusion
XX protein or a multimeric protein and at least one additional non-ixodes
XX scapularis polypeptide, can be administered to a subject to confer tick
XX immunity and therefore prevent infection by a tick-borne disease.
XX Inhibiting coagulation factor Xa activity comprises administering to a
XX subject a polypeptide selected from Salp14A, Salp9A or a fragment of
XX these having Xa inhibiting activity, inhibiting histamine activity
XX comprises administering a Salp25D polypeptide or its histamine binding
XX fragment to a subject, and inhibiting or preventing an inflammatory
XX response comprises administering a polypeptide selected from a Salp15,
XX Salp25C, Salp13 or a fragment of these having the same activity.
XX The protein or its immunogenic fragment may be used to produce
XX monoclonal antibodies that are screened for their ability to confer
XX tick immunity when used to immunise naive animals. The antibody can be
XX used to screen for expression of I. scapularis polypeptides, either in
XX libraries constructed from I. scapularis nucleic acid molecules or from
XX other samples in which proteins may be present, and for purifying or
XX removing polypeptides from a given sample to block or bind to specific
XX epitopes on the polypeptides and to direct various molecules, such as
XX toxins, to ticks.
XX Sequence 222 AA:
SQ
Query Match 2.8%; Score 7; DB 22; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 VLLCGVL 14
Db 9 VLLCGVL 15
RESULT 44
AAW27087
ID AAW27087 standard; Protein; 229 AA.
XX
XX AAW27087;
XX
XX 28-JAN-1998 (first entry)
XX Human transforming growth factor alpha HIII.
XX human transforming growth factor; TGF; TGF-alpha-HIII; angiogenesis;
XX embryogenesis; ocular disorder; kidney disorder; liver disorder;
XX neuronal disorder; alopecia; inflammation.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..25
XX /label= signal_peptide
XX /note= "putative"
XX Protein 26..229
XX /label= mature_protein
XX Region 1..177
/label= soluble_portion_of_protein
126..177
/label= active_site
178..204
/label= transmembrane_portion
/note= "putative"
W09725349-A1.
17-JUL-1997.
04-JAN-1996; 96WO-US00149.
04-JAN-1996; 96WO-US00149.
(HUMA-) HUMAN GENOME SCI INC.
Wei Y;
WPI; 1997-372817/34.
DR N-PSDB; AAT85082.
XX New human transforming growth factor-alpha homologue - used for
XX developing products for treating e.g. neurological disorders, kidney
XX and liver disorders, tumours, wounds, hair loss or skin disorders
XX Claim 15; Page 47; 63pp; English.
XX This protein has been putatively identified as a human transforming
XX growth factor (TGF) alpha analogue, TGF-alpha-HIII. The protein can
XX stimulate angiogenesis, embryogenesis, cell differentiation and function.
XX It can be used for therapeutic purposes for restoration or enhancement of
XX neurological functions diminished as a result of trauma or other damaging
XX pathologies such as AIDS dementia and senile dementia, to treat ocular
XX disorders, e.g. corneal inflammation, to destroy target cells, to treat
XX tumours, kidney or liver disorders or to treat wounds, burns or ulcers.
XX The polypeptide can also be used in the modulation of angiogenesis, bone
XX resorption, immune response, and synaptic and neuronal effector
XX functions, or the arachidonic acid cascade. It can also be used in
XX applications related to terminal differentiation e.g. in
XX hyperproliferative disorders such as inflammation or psoriasis and for
XX alopecia, hair loss or other skin conditions which affect hair follicular
XX development. Antagonists to TGF-alpha-HIII can be used for treating
XX tumours or skin disorders such as psoriasis. The products can also
XX be used for diagnosis and detection of the above disorders.
XX Sequence 229 AA:
SQ
Query Match 2.8%; Score 7; DB 18; Length 229;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 243 TQLOTLII 249
Db 117 TQLTLII 123
RESULT 45
AAV13944
ID AAV13944 standard; Protein; 229 AA.
XX
XX AAV13944;
XX
XX 14-JUL-1999 (first entry)
XX Human transmembrane protein, HP10435.
XX
XX Transmembrane protein; human; cell membrane; proliferation; diagnosis;
XX cell differentiation; carcinostatic agent; probe; gene therapy;
XX signal transduction; apoptosis; inhibitor;
XX phosphatidylethanolamine N-methyltransferase.
XX

```

OS Homo sapiens.
 XX W09918203-A2.
 PN
 XX
 XX 15-APR-1999.
 PD
 XX
 XX 05-OCT-1998; 98WO-JPU4475.
 PF
 XX
 XX 08-OCT-1997; 97JP-0276271.
 PR
 XX (PROT-) PROTEGENE, INC.
 PA (SAGA) SAGAMI CHEM RES CENT.
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 XX WPI: 1999-277268/23.
 XX N-PSDB; AAX36812, AAX36813.
 DR
 DR Human transmembrane proteins and nucleotide sequences
 XX
 PT Claim 1; Page 90-91; 139pp; English.
 PS
 XX This sequence is a human transmembrane protein of the invention.
 CC All of the proteins exist in the cell membrane, so are considered to be
 CC proteins controlling the proliferation and differentiation of the cells.
 CC They may be useful as carcinostatic agents or as antigens for preparing
 CC antibodies against the proteins. The cDNAs can be used as probes for
 CC gene diagnosis and gene sources for gene therapy, as well as for
 CC large-scale expression of the proteins. The HPO1498 (see AAY13939)
 CC protein may be associated with signal transduction associated with
 CC apoptosis, and therefore useful in inhibition of apoptosis. The HPO1962
 CC (see AAY13943) protein can be used to treat diseases associated with
 CC phosphatidylethanolamine N-methyltransferase. The proteins are
 CC identified by the presence of a hydrophobic transmembrane region,
 CC knowledge of the protein function is not required, as in e.g. methods of
 CC expression cloning.
 XX
 SQ Sequence 229 AA;
 Query Match 2.8%; Score 7; DR 20; Length 229;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 243 TQLQTLI 249
 Db 117 tqqltll 123
 RESULT 46
 AAY05282
 ID AAY05282 standard; Protein: 229 AA.
 XX
 XX AAY05282;
 AC
 XX 22-JUN-1999 (first entry)
 DT
 XX EGF-like homologue PRO240.
 DE
 XX Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO261; PRO246;
 KW EBAF-2; inhibitor; tumour growth; cancer; EGF-like homologue;
 KW FGF-8 homologue.
 XX
 OS Homo sapiens.
 XX W09914327-A2.
 PN
 XX 25-MAR-1999.
 PD
 XX 10-SEP-1998; 98WO-US18824.
 PF
 XX 25-NOV-1997; 97US-0066840.
 PR
 PR 17-SEP-1997; 97US-0059114.

PR 17-SEP-1997; 97US-0059117.
 PR 18-SEP-1997; 97US-0059263.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 24-OCT-1997; 97US-0062816.
 PR 29-OCT-1997; 97US-0063704.
 XX
 XX (GETH) GENENTECH INC.
 PA Botstein D, Goddard A, Gurney A, Hillan K, Lawrence DA;
 PI Roy M, Wood WJ;
 XX
 XX WPI: 1999-229532/19.
 DR N-PSDB; AAX28432.
 XX
 XX Antibodies against specific proteins overexpressed in tumours
 PT
 XX Example 1; Fig 12; 130pp; English.
 PS
 XX This sequence represents the EGF-like homologue PRO240.
 CC The invention relates to antibodies (Ab) that bind to any of the
 CC polypeptides (I) designated PRO187; PRO533; PRO214; PRO240; PRO211;
 CC PRO230; PRO261; PRO246 or EBAF-2. The Ab, or other agents that inhibit
 CC expression and/or activity of (I) are used: (i) to inhibit growth of
 CC tumours; and (ii) as diagnostic/prognostic reagents for detection of
 CC quantification of (I) in cells or tissues, by standard immunoassays, with
 CC overexpression being indicative of cancer. For therapeutic use, the Ab
 CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.
 CC Genes expressing (I), many of which are growth factor homologues, are
 CC overexpressed in some cases of cancer.
 XX
 SQ Sequence 229 AA;
 Query Match 2.8%; Score 7; DB 20; Length 229;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 243 TQLQTLI 249
 Db 117 tqqltll 123
 RESULT 47
 AAB33419
 ID AAB33419 standard; Protein: 229 AA.
 XX
 XX AAB33419;
 AC
 XX 29-JAN-2001 (first entry)
 DT
 XX Human PRO240 protein UNQ214 SEQ ID NO:26.
 DE
 XX Human; immune related disease; diagnosis; antiinflammatory; cardiant;
 KW dermatological; antiarthritic; antirheumatic; immunosuppressive;
 KW haemostatic; antithyroid; antidiabetic; nootropic; neuroprotective;
 KW antianaemic; hepatotropic; virucide; antipsoriatic; antiallergic;
 KW antiasthmatic; systemic lupus erythematosus; rheumatoid arthritis;
 KW osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;
 KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;
 KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;
 KW autoimmune thrombocytopaenia; immune-mediated renal disease;
 KW demyelinating disease; hepatobiliary disease; Whipple's disease;
 KW inflammatory bowel disease; gluten-sensitive enteropathy;
 KW autoimmune disease; immune-mediated skin disease; allergic disease;
 KW immunological disease; transplantation associated disease;
 KW graft rejection; graft-versus-host-disease.
 XX
 OS Homo sapiens.
 XX W0200053758-A2.
 PN
 XX

PD 14-SEP-2000.
 XX 02-MAR-2000; 2000WO-US05841.
 PF 08-MAR-1999; 99WO-US05028.
 XX 10-MAR-1999; 99US-0123618.
 PR 12-MAR-1999; 99US-0123957.
 PR 23-MAR-1999; 99US-0125775.
 PR 12-APR-1999; 99US-0128849.
 PR 20-APR-1999; 99WO-US08615.
 PR 28-APR-1999; 99US-0131445.
 PR 04-MAY-1999; 99US-0132371.
 PR 14-MAY-1999; 99WO-US12252.
 PR 02-JUN-1999; 99US-0141037.
 PR 23-JUN-1999; 99US-0144758.
 PR 20-JUL-1999; 99US-0145698.
 PR 26-JUL-1999; 99US-0146222.
 PR 28-JUL-1999; 99WO-US20111.
 PR 01-SEP-1999; 99WO-US20594.
 PR 08-SEP-1999; 99WO-US20944.
 PR 13-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 15-SEP-1999; 99WO-US23089.
 PR 05-OCT-1999; 99US-0162806.
 PR 29-OCT-1999; 99WO-US28214.
 PR 29-NOV-1999; 99WO-US28313.
 PR 30-NOV-1999; 99WO-US28409.
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30999.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 XX
 PA (GETH) GENENTECH INC.
 PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;
 PI Kabakoff RC, Lu Y, Fan J, Pennica D, Shelton DL, Smith V;
 PI Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;
 XX
 DR WPI; 2000-572271/53.
 DR N-PSDB; AAC58584.
 XX
 PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of
 PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid
 PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -
 XX
 PS Claim 33; Fig 12; 309pp; English.
 XX
 CC The present invention describes sixty four human PRO proteins which can
 CC be used in the treatment of immune related diseases. The human PRO
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for
 CC treating and diagnosing immune related disorders. The disorders are
 CC selected from systemic lupus erythematosus, rheumatoid arthritis,
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthritis,
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic
 CC anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,
 CC immune-mediated renal disease, demyelinating diseases of the central
 CC and peripheral nervous systems, hepatobiliary diseases, inflammatory
 CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,
 CC autoimmune or immune-mediated skin diseases, allergic diseases,
 CC immunological diseases of the lung, and transplantation associated

CC diseases including graft rejection and graft-versus-host-disease.
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and
 CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.
 XX
 SQ Sequence 229 AA;
 Query Match 2.8%; Score 7; DB 21; Length 229;
 Best Local Similarity 100.0%; Pred. No. 2.6e-02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 243 TQLQTLI 249
 DQ 117 tq|qtli 123
 |||||
 RESULT 48
 AAY88570
 ID AAY88570 standard; Protein: 229 AA.
 XX
 AC AAY88570;
 XX
 DT 09-AUG-2000 (first entry)
 XX
 DE Human PRO240 amino acid sequence.
 XX
 KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;
 KW PRO317; tumour growth inhibitor; cancer; diagnosis; treatment; human;
 KW cell growth proliferation; serrate precursor; C-serrate-1; ADPPT;
 KW antibody dependent enzyme mediated prodrug therapy; chromosome 2.
 XX
 OS Homo sapiens.
 XX
 PN WO200015666-A2.
 XX
 PD 23-MAR-2000.
 XX
 PF 08-SEP-1999; 99WO-US20594.
 PR 10-SEP-1998; 98US-0099803.
 PR 10-SEP-1998; 98WO-US18824.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI, Botstein D;
 XX
 DR WPI; 2000-271386/23.
 DR N-PSDB; AAA30036.
 XX
 PT New isolated antibodies which bind to specific polypeptides used for
 PT diagnosis and treatment of neoplastic cell growth and proliferation -
 XX
 PS Example 4; Fig 8; 200pp; English.
 XX
 CC This sequence represents a human PRO240 amino acid sequence. PRO240
 CC shares sequence homology with the D. melanogaster serrate precursor
 CC protein and the Gallus gallus C-serrate-1 protein. The PRO240 gene is
 CC located on chromosome 2. The invention relates to isolated antibodies
 CC which bind to a polypeptide. The "PRO" polypeptides are encoded by genes
 CC which are over expressed in the genome of tumour cells. Vectors and host
 CC cells comprising the nucleic acid encoding the antibodies are used in the
 CC production of the antibodies. The antibodies and nucleic acids encoding
 CC them are used for diagnosing a tumour in a mammal. The antibodies are
 CC used for inhibiting the growth of tumour cells and identifying compounds
 CC that inhibit a biological or immunological activity of and/or expression
 CC of a PRO187, PRO533, PRO214, PRO240, PRO211, PRO230, PRO261, PRO246 or
 CC PRO317 polypeptide. The antibody can be used in antibody dependent enzyme
 CC mediated prodrug therapy (ADPPT) by conjugating the antibody to a
 CC prodrug-activating enzyme which converts a prodrug to an anti-cancer
 CC drug. The antibodies can be fluorescently labelled and monitored by light
 CC microscopy, flow cytometry or fluorimetry for diagnosis and prognosis of

CC tumours.
 XX Sequence 229 AA;
 SQ

Query Match 2.8%; Score 7; DB 21; Length 229;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 243 TQIQTLI 249
 Db 117 tqiqtl 123
 |||||

RESULT 49
 AAU04295
 ID AAU04295 standard; Protein: 229 AA.
 XX
 AC AAU04295;
 XX
 XX 24-OCT-2001 (first entry)
 DT
 XX Transforming growth factor (TGF) alpha HIII.
 DE
 XX Human; TGF alpha HIII; transforming growth factor alpha HIII; cancer;
 KW diagnostic; therapeutic; immune disorder; multiple sclerosis; HIV;
 KW systemic lupus erythematosus; human immuno-deficiency virus; HIV;
 KW hyperproliferative disorder; Gaucher's disease; cardiovascular disease;
 KW Scimitar syndrome; Chaga's cardiomyopathy; coronary arteriosclerosis;
 KW angiogenic disorder; corneal graft; neovascularisation; wound healing;
 KW diabetic retinopathy; neurological disorder; Huntington's chorea;
 KW Alzheimer's disease; Parkinson's disease.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..25
 FT /label= Signal_peptide
 FT Protein 26..229 "Mature TGF alpha HIII"
 FT /note= "Mature TGF alpha HIII"
 FT Active-site 126..177
 FT Region 178..204
 FT /note= "Transmembrane region"
 XX
 XX WO200140251-A1.
 PN
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000WO-US32745.
 XX
 PR 02-DEC-1999; 99US-0168387.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 PI Wei Y;
 XX
 XX WPI; 2001-441480/47.
 DR N-PSDB; AAS08543.
 XX
 XX Nucleic acid encoding human transforming growth factor alpha III
 XX (TGFA), useful for preventing, diagnosing and/or treating e.g. Cancer
 XX and Parkinson's disease
 XX
 XX Claim 11; Fig 1; 302pp; English.
 PS
 XX The sequence represents the amino acid sequence of human transforming
 XX growth factor (TGF) alpha HIII. TGF alpha HIII nucleic acid and protein
 XX may be used in the prevention, diagnosis and treatment of diseases
 XX associated with inappropriate polypeptide expression, for example immune
 XX disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
 XX human immuno-deficiency virus (HIV) infections), hyperproliferative
 XX disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
 XX (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary

CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC infectious diseases and/or for promoting wound healing, regeneration
 CC and/or chemotaxis (full details given in specification). Additionally,
 CC the nucleic acid may be used to produce the secreted polypeptides, by
 CC inserting the nucleic acids into a host cell and culturing the cell to
 CC express the protein. It may also be used as a DNA probe in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples, and therefore which patients may be in need of
 CC restorative therapy. The polypeptides may also be used as antigens in the
 CC production of antibodies against TGF alpha HIII and in assays to identify
 CC modulators of TGF alpha HIII. The anti-TGF alpha HIII antibodies may also
 CC be used as diagnostic agents for detecting the presence of TGF alpha HIII
 CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)).
 XX
 SQ Sequence 229 AA;

Query Match 2.8%; Score 7; DB 22; Length 229;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 243 TQIQTLI 249
 Db 117 tqiqtl 123
 |||||

RESULT 50
 AAB20112
 ID AAB20112 standard; Protein: 229 AA.
 XX
 AC AAB20112;
 XX
 XX 30-APR-2001 (first entry)
 DT
 XX Human immunostimulant PRO240.
 DE
 XX
 KW PRO240; UNQ214; human; immune disease; autoimmune disease;
 KW antirheumatic; antiarthritic; antiinflammatory; antianaemic;
 KW immunosuppressive; antithyroid; antidiabetic; neuroprotective;
 KW hepatotropic; virucide; dermatological; antipsoriatic;
 KW antiasthmatic; antiallergic; immunostimulant; serrate; lung cancer.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..30
 FT /label= Signal_peptide
 FT Protein 31..229
 FT /label= Mature_protein
 FT Domain 198..213
 FT /note= "transmembrane domain"
 FT Region 181..193
 FT /note= "epidermal growth factor domain cysteine
 FT pattern signature"
 FT Modified-site 44..48
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 79..83
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 157..161
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 168..172
 FT /note= "Asn is N-glycosylated"
 FT Modified-site 7..13
 FT /note= "N-myristoylation site"
 FT Modified-site 24..30
 FT /note= "N-myristoylation site"
 FT Modified-site 40..46
 FT /note= "N-myristoylation site"
 FT Modified-site 70..76
 FT /note= "N-myristoylation site"
 FT Modified-site 132..138

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 16:39:58 : Search time 13.09 seconds
(without alignments)
464,627 Million cell updates/sec

Title: US-10-020-139-2

Perfect score: 249

Sequence: 1 MLQLKWLKLLCCVLTGTS.....NVIQVVDNPGHKTQLTLI 249

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database :

Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	3.2	561	US-08-679-635A-7	Sequence 7, Appli
2	7	2.8	523	US-09-550-338-2	Sequence 2, Appli
3	7	2.8	664	US-09-295-186-17	Sequence 17, Appli
4	7	2.8	777	US-08-477-396A-4	Sequence 4, Appli
5	7	2.8	866	US-08-483-101-4	Sequence 5, Appli
6	7	2.8	1048	US-09-356-952-5	Sequence 5, Appli
7	6	2.4	12	US-08-250-789A-26	Sequence 26, Appli
8	6	2.4	12	US-08-250-789A-62	Sequence 62, Appli
9	6	2.4	14	US-08-300-928C-23	Sequence 23, Appli
10	6	2.4	14	US-08-430-944D-23	Sequence 23, Appli
11	6	2.4	14	US-08-430-014-23	Sequence 23, Appli
12	6	2.4	14	US-08-431-184-23	Sequence 23, Appli
13	6	2.4	14	PCT-US93-02462-13	Sequence 13, Appli
14	6	2.4	16	US-08-300-928C-22	Sequence 22, Appli
15	6	2.4	16	US-08-430-944D-22	Sequence 22, Appli
16	6	2.4	16	US-08-430-014-22	Sequence 22, Appli
17	6	2.4	16	US-08-431-184-22	Sequence 22, Appli
18	6	2.4	16	PCT-US93-02462-12	Sequence 12, Appli
19	6	2.4	19	US-07-807-529A-10	Sequence 10, Appli
20	6	2.4	19	US-08-300-928C-20	Sequence 20, Appli
21	6	2.4	19	US-08-430-944D-20	Sequence 20, Appli
22	6	2.4	19	US-08-430-014-20	Sequence 20, Appli
23	6	2.4	19	US-08-431-184-20	Sequence 20, Appli
24	6	2.4	19	PCT-US93-02462-10	Sequence 10, Appli
25	6	2.4	20	US-07-807-529A-29	Sequence 29, Appli
26	6	2.4	20	US-08-300-928C-78	Sequence 78, Appli
27	6	2.4	20	US-08-430-944D-78	Sequence 78, Appli

28	6	2.4	20	3	US-08-430-014-78	Sequence 78, Appli
29	6	2.4	20	3	US-08-431-184-78	Sequence 78, Appli
30	6	2.4	20	5	PCT-US93-02462-14	Sequence 14, Appli
31	6	2.4	22	1	US-08-383-753-7	Sequence 7, Appli
32	6	2.4	22	2	US-08-586-772-7	Sequence 7, Appli
33	6	2.4	22	2	US-08-959-512-7	Sequence 7, Appli
34	6	2.4	22	4	US-09-512-983-7	Sequence 7, Appli
35	6	2.4	35	3	US-08-782-997A-24	Sequence 24, Appli
36	6	2.4	38	2	US-08-807-332B-6	Sequence 6, Appli
37	6	2.4	38	4	US-09-338-876-6	Sequence 6, Appli
38	6	2.4	70	1	US-07-662-193-3	Sequence 3, Appli
39	6	2.4	75	6	5252466-3	Patent No. 5252466
40	6	2.4	86	2	US-08-248-839C-67	Sequence 67, Appli
41	6	2.4	86	2	5459046-11	Patent No. 5459046
42	6	2.4	85	3	US-08-855-531D-9	Sequence 9, Appli
43	6	2.4	88	1	US-08-855-526B-9	Sequence 9, Appli
44	6	2.4	88	1	US-07-807-529A-4	Sequence 4, Appli
45	6	2.4	88	5	PCT-US93-02462-4	Sequence 4, Appli
46	6	2.4	91	2	US-08-997-080-143	Sequence 143, App
47	6	2.4	91	2	US-08-997-362-143	Sequence 143, App
48	6	2.4	91	4	US-09-095-855-143	Sequence 143, App
49	6	2.4	91	4	US-09-324-542-143	Sequence 143, App
50	6	2.4	92	1	US-07-807-529A-2	Sequence 2, Appli
51	6	2.4	92	5	PCT-US93-02462-2	Sequence 2, Appli
52	6	2.4	94	1	US-07-662-193-1	Sequence 1, Appli
53	6	2.4	94	3	US-08-300-928C-2	Sequence 2, Appli
54	6	2.4	94	3	US-08-300-928C-11	Sequence 11, Appli
55	6	2.4	94	3	US-08-430-944D-2	Sequence 2, Appli
56	6	2.4	94	3	US-08-430-944D-11	Sequence 11, Appli
57	6	2.4	94	3	US-08-430-014-2	Sequence 2, Appli
58	6	2.4	94	3	US-08-430-014-11	Sequence 11, Appli
59	6	2.4	94	3	US-08-431-184-2	Sequence 2, Appli
60	6	2.4	94	3	US-08-431-184-11	Sequence 11, Appli
61	6	2.4	94	4	US-09-142-078-58	Sequence 58, Appli
62	6	2.4	94	4	US-09-357-141-58	Sequence 58, Appli
63	6	2.4	96	1	US-07-662-193-2	Sequence 2, Appli
64	6	2.4	96	3	US-08-300-928C-4	Sequence 4, Appli
65	6	2.4	96	3	US-08-300-928C-12	Sequence 12, Appli
66	6	2.4	96	3	US-08-430-944D-4	Sequence 4, Appli
67	6	2.4	96	3	US-08-430-944D-12	Sequence 12, Appli
68	6	2.4	96	3	US-08-430-014-4	Sequence 4, Appli
69	6	2.4	96	3	US-08-430-014-12	Sequence 12, Appli
70	6	2.4	96	3	US-08-431-184-4	Sequence 4, Appli
71	6	2.4	96	3	US-08-431-184-12	Sequence 12, Appli
72	6	2.4	98	4	US-09-142-078-54	Sequence 54, Appli
73	6	2.4	98	4	US-09-357-141-54	Sequence 54, Appli
74	6	2.4	98	6	5459046-6	Patent No. 5459046
75	6	2.4	99	1	US-08-202-389-25	Sequence 25, Appli

ALIGNMENTS

RESULT 1
US-08-679-635A-7
: Sequence 7, Application US/08679635A
: Patent No. 5985643
: GENERAL INFORMATION:
: APPLICANT: Tomasz, Alexander
: APPLICANT: Delencastre, Herminia
: TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
: TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David A. Jackson, Esq.
: STREET: 411 Hackensack Ave, Continental Plaza, 4th
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/679,635A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-679-635A-7

```

```

Query Match      3.2%; Score 8; DB 2; Length 561;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 58 LKVDLGLV 65
   |||||
DB 162 LKVDLGLV 169

```

```

RESULT 2
US-09-550-338-2
; Sequence 2, Application US/09550338
; Patent No. 6210951
; GENERAL INFORMATION:
; APPLICANT: KAWASAKI, Hisashi
; APPLICANT: USUDA, Yoshihiro
; APPLICANT: MIHARA, Yoshihiro
; APPLICANT: KURAHASHI, Osamu
; TITLE OF INVENTION: GMP Synthetase and Gene Coding for the Same
; FILE REFERENCE: 0010-1101-0
; CURRENT APPLICATION NUMBER: US/09/550,338
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: JP 11-114787
; PRIOR FILING DATE: 1990-04-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum (Corney)bacterium glutamicum)
US-09-550-338-2

```

```

Query Match      2.8%; Score 7; DB 4; Length 523;
Best Local Similarity 100.0%; Pred. No. 1.6e-02;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 39 VLHGLE 45
   |||||
DB 119 VLHGLE 125

```

```

RESULT 3
US-09-295-186-17
; Sequence 17, Application US/09295186B
; Patent No. 6127137
; GENERAL INFORMATION:
; APPLICANT: Hasida, Miyoko

```

```

; APPLICANT: Tsutsumi, No. 6127137iko
; APPLICANT: Halkier, Torben
; APPLICANT: Stringer, Mary Ann
; TITLE OF INVENTION: An Acidic Phospholipase, Production, and
; FILE REFERENCE: 4953.204-US
; CURRENT APPLICATION NUMBER: US/09/295,186B
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: 1215/96
; PRIOR FILING DATE: 1996-10-31
; PRIOR APPLICATION NUMBER: PCT/DK97/00490
; PRIOR FILING DATE: 1997-10-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 664
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-295-186-17

```

```

Query Match      2.8%; Score 7; DB 3; Length 664;
Best Local Similarity 100.0%; Pred. No. 2e-02;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 122 GNLSP 128
   |||||
DB 457 GNLSP 463

```

```

RESULT 4
US-08-477-396A-4
; Sequence 4, Application US/08477396A
; Patent No. 5872235
; GENERAL INFORMATION:
; APPLICANT: Chen, Lan Bo
; APPLICANT: Bao, Shideng
; APPLICANT: Liu, Xuan
; TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
; TITLE OF INVENTION: ISOLATING SAME
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,396A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/146,488
; FILING DATE: 29-OCT-1993
; APPLICATION NUMBER: US 08/448,388
; FILING DATE: 28-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12502
; FILING DATE: 31-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Heine, Holliday C.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DFCI-333BX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 4:

```

SEQUENCE CHARACTERISTICS:
 LENGTH: 777 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-477-396A-4

Query Match 2.8%; Score 7; DB 2; Length 777;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 VLKSSA 70
 Db 33 VLKSSA 39

RESULT 5
 US-08-483-101-4
 : Sequence 4, Application US/08483101
 : Patent No. 5932715

GENERAL INFORMATION:
 : APPLICANT: Scott, June R.
 : APPLICANT: Froehlich, Barbara
 : APPLICANT: Caron, Judy
 : TITLE OF INVENTION: CS2 Proteins and Coding Sequences
 : NUMBER OF SEQUENCES: 17
 : CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee and Winner, P.C.
 STREET: 5370 Manhattan Circle, Suite 201
 CITY: Boulder
 STATE: Colorado
 COUNTRY: US
 ZIP: 80303

COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/483,101
 : FILING DATE: 07-JUN-1995

CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 : NAME: Feiber, Donna M.
 : REGISTRATION NUMBER: 33878
 : REFERENCE/DOCKET NUMBER: 6-95
 : TELEPHONE: (303) 499-8080
 : TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 866 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 US-08-483-101-4

Query Match 2.8%; Score 7; DB 2; Length 866;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 VTANVT 135
 Db 761 VTANVT 767

RESULT 6
 US-09-356-952-5
 : Sequence 5, Application US/09356952
 : Patent No. 6117663
 : GENERAL INFORMATION:

APPLICANT: Boriack-Sjodin, Ann
 APPLICANT: Margarit, S. M.
 APPLICANT: Bor-Sogli, Dafna
 APPLICANT: Cole, Philip
 APPLICANT: Kuriyan, John
 : TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
 : FILE REFERENCE: 600-1-228N
 : CURRENT APPLICATION NUMBER: US/09/356,952
 : CURRENT FILING DATE: 1999-07-19
 : EARLIER APPLICATION NUMBER: 60/093,631
 : EARLIER FILING DATE: 1998-07-21
 : NUMBER OF SEQ ID NOS: 14
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 5
 : LENGTH: 1048
 : TYPE: PRT
 : ORGANISM: Saccharomyces cerevisiae
 US-09-356-952-5

Query Match 2.8%; Score 7; DB 3; Length 1048;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 SLDLLTA 155
 Db 49 SLDLLTA 55

RESULT 7
 US-08-250-789A-26
 : Sequence 26, Application US/08250789A
 : Patent No. 5635597
 : GENERAL INFORMATION:
 : APPLICANT: Barrett, Ronald W.
 : APPLICANT: Chernov-Rogan, Tania
 : APPLICANT: Davis, Ann M.
 : TITLE OF INVENTION: Peptides That Bind to IL-2 Receptors
 : NUMBER OF SEQUENCES: 194
 : CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend Kourile and Crew
 STREET: One Market Plaza, Steuart Tower, Suite 2000
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94105

COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/250,789A
 : FILING DATE: 27-MAY-1994
 : CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
 : NAME: No. 5635597v1el, Vernon A.
 : REGISTRATION NUMBER: 32,483
 : REFERENCE/DOCKET NUMBER: 16528A-57/1043
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 415-326-2400
 : TELEFAX: 415-326-2422
 : INFORMATION FOR SEQ ID NO: 26:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 12 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: peptide
 US-08-250-789A-26

Query Match 2.4% Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 VAVLGE 173
Db 5 VAVLGE 10

RESULT 8
US-08-250-789A-62
Sequence 62, Application US/08250789A
Patent No. 5635597
GENERAL INFORMATION:
APPLICANT: Barlett, Ronald W.
APPLICANT: Chernov-Rodan, Tania
APPLICANT: Davis, Ann M.
TITLE OF INVENTION: Peptides That Bind to IL-2 Receptors
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/250,789A
FILING DATE: 27-MAY-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: No. 5635597v1el, Vernon A.
REGISTRATION NUMBER: 32,483
REFERENCE/DOCKET NUMBER: 16528A-57/1043
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-250-789A-62

Query Match 2.4% Score 6; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 VAVLGE 173
Db 5 VAVLGE 10

RESULT 9
US-08-300-928C-23
Sequence 23, Application US/08300928C
Patent No. 6019972
GENERAL INFORMATION:
APPLICANT: Geffer, Malcolm L. et al.
TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
TITLE OF INVENTION: PROTEIN (TRFP)
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET

CITY: WALTHAM
STATE: Massachusetts
COUNTRY: USA
ZIP: 02145
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/300,928C
FILING DATE: September 2, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,529
FILING DATE: December 13, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: AMY E. MANDRAGOURAS
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: 002.6US(IMI-044)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-300-928C-23

Query Match 2.4% Score 6; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLLOK 188
Db 3 LSLLOK 8

RESULT 10
US-08-430-944D-23
Sequence 23, Application US/08430944D
Patent No. 6025162
GENERAL INFORMATION:
APPLICANT: Bruce L. Rogers et al.
TITLE OF INVENTION: A HUMAN T CELL REACTIVE FELINE PROTEIN
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,944D
FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/430,014
FILING DATE: 27-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,928
FILING DATE: 02-SEPT-1994

```

;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Amy E. Mandragouras
;
; REGISTRATION NUMBER: 36,207
;
; REFERENCE/DOCKET NUMBER: JMI-044DV2
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (617)227-7400
;
; TELEFAX: (617)742-4214
;
; INFORMATION FOR SEQ ID NO: 23:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 14 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: peptide
;
; FRAGMENT TYPE: internal
;
; US-08-430-944D-23

```

Query Match 2.4%; Score 6; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels

Qy 183 LSLDK 188
11111
Db 3 LSLDK 8

```

RESULT 11
US-08-430-014-23
: Sequence 23, Application US/08430014
: Patent NO. 6048962
: GENERAL INFORMATION:
: APPLICANT: GETTER, Malcolm L. et al.
: TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
: NUMBER OF SEQUENCES: 101
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
: STREET: 610 LINCOLN STREET
: CITY: WALTHAM
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02145
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/430,014
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/300,928
: FILING DATE:

```

Query Match 2.48; Score 6; DB 3; Length 14;

```

Best Local Similarity 100.0%; Pred. NO. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDDK 188
      |||||
Db 3 LSLDDK 8

RESULT 12
US-08-431-184-23
; Sequence 23, Application US/08431184
; Patent No. 6120769
; GENERAL INFORMATION:
; APPLICANT: Bruce L. Rogers et al.
; TITLE OF INVENTION: A HUMAN T CELL REACTIVE FELINE PROTEIN
; TITLE OF INVENTION:

```

```

Query Match      2.4%; Score 6; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels

```

Qy	183	LSLKD	188
Db	3	LSLKD	8

```

RESULT      13
PCT-US93-02462-13
: Sequence 13, Application PC/TUS9302462
:
: GENERAL INFORMATION:
: APPLICANT: Gefter, Malcolm L.
: APPLICANT: Garman, Richard D.
: APPLICANT: Greenstein, Julia L.
: APPLICANT: Kuo, Mei-chang
: APPLICANT: Briner, Thomas J.

```



```

1  APPLICANT: Morville, Malcolm
2  TITLE OF INVENTION: PEPTIDES USEFUL FOR TOLERIZATION
3  NUMBER OF SEQUENCES: 14
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: LAHIVE & COCKFIELD
6  STREET: 60 State Street
7  CITY: Boston
8  STATE: MA
9  COUNTRY: USA
10 ZIP: 02109
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: ASCII TEXT
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: PCT/US93/02462
18 FILING DATE: 19930325
19 CLASSIFICATION:
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: US 08/006,116
22 FILING DATE: 15-JAN-1993
23 APPLICATION NUMBER: US 07/884,718
24 FILING DATE: 15-MAY-1992
25 APPLICATION NUMBER: 07/857,311
26 FILING DATE: 25-MAR-1992
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Mandragoras, Amy E.
29 REGISTRATION NUMBER: 36,257
30 REFERENCE/DOCKET NUMBER: IPC-031PC
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (617) 227-7400
33 INFORMATION FOR SEQ ID NO: 13:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 14 amino acids
36 TYPE: AMINO ACID
37 TOPOLOGY: linear
38 MOLECULE TYPE: peptide
39 FRAGMENT TYPE: internal
40 PCT-US93-02462-13

```

APPLICATION NUMBER: US/08/300,928C
 FILING DATE: September 2, 1994
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 07/807,529
 FILING DATE: December 13, 1991
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: AMY E. MANDRAGOURAS
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: 002.60US(IMI-044)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 22:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 16 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal
 US-08-300-928C-22

```

Query Match          2.4%; Score 6; DB 5; Length 14;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY      183 LSLDDK 188
      HHHH
Db       3 LSLDDK 8

RESULT 14
US-08-300-928C-22
; Sequence 22, Application US/08300928C
; Patent No 6019972
; GENERAL INFORMATION:
; APPLICANT: GEFTER, Malcolm L. et al.
; TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
; TITLE OF INVENTION: PROTEIN (TRFP)
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
; STREET: 610 LINCOLN STREET
; CITY: WALTHAM
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02145
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:

```

NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,944D
FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/430,014
FILING DATE: 27-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,928
FILING DATE: 02-SEPT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-044DV2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-430-9440-22

Query Match 2.4%; Score 6; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
Db 8 LSLDK 13

RESULT 16

US-08-430-014-22
Sequence 22, Application US/08430014
Patent No. 6048962
GENERAL INFORMATION:
APPLICANT: GEFTER, Malcolm L. et al.
TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
TITLE OF INVENTION: PROTEIN (TRFP)
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: Massachusetts
COUNTRY: USA
ZIP: 02145

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,014
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/300,928
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: AMY E. MANDRAGOURAS
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: 002.6US(IMI-044)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-430-014-22

Query Match 2.4%; Score 6; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
Db 8 LSLDK 13

RESULT 17

US-08-431-184-22

Sequence 22, Application US/08431184
Patent No. 6120769
GENERAL INFORMATION:
APPLICANT: Bruce L. Rogers et al.
TITLE OF INVENTION: A HUMAN T CELL REACTIVE FELINE PROTEIN
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,184
FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/430,014
FILING DATE: 27-APR-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,928
FILING DATE: 02-SEPT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-044DV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-431-184-22

Query Match 2.4%; Score 6; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
Db 8 LSLDK 13

RESULT 18

PCT-US93-02462-12
Sequence 12, Application PC/TUS9302462
GENERAL INFORMATION:
APPLICANT: Gefter, Malcolm L.
APPLICANT: Garman, Richard D.
APPLICANT: Greenstein, Julia L.
APPLICANT: Kuo, Mei-chang
APPLICANT: Briner, Thomas J.
APPLICANT: Morville, Malcolm
TITLE OF INVENTION: PEPTIDES USEFUL FOR TOLERIZATION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02462
; FILING DATE: 19930325
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/006,116
; FILING DATE: 15-JAN-1993
; APPLICATION NUMBER: US 07/884,718
; FILING DATE: 15-MAY-1992
; APPLICATION NUMBER: 07/857,311
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-03LPC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
;
; PCT-US93-02462-12

```

```

Query Match 2.4%; Score 6; DB 5; Length 16;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 183 LSLLDK 188
      |||||
Db 8 LSLLDK 13

```

```

RESULT 19
US-07-807-529A-10
; Sequence 10, Application US/07807529A
; Patent No. 5547669
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce L.
; APPLICANT: Morgenstern, Jay
; APPLICANT: Bond, Julian F.
; APPLICANT: Garman, Richard D.
; APPLICANT: Greenstein, Julia L.
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Morville, Malcolm
; TITLE OF INVENTION: RECOMBITOPE PEPTIDES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
; STREET: One Kendall Square, Building 600
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02139
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,529A
; FILING DATE: 19911213
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,276

```

```

; FILING DATE: 28-FEB-1991
; APPLICATION NUMBER: US 07/431,565
; FILING DATE: 03-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Channing, Stacey L.
; REGISTRATION NUMBER: 31,095
; REFERENCE/DOCKET NUMBER: IPC-027/Imi-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 434-0060
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
;
; US-07-807-529A-10

```

```

Query Match 2.4%; Score 6; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 183 LSLLDK 188
      |||||
Db 8 LSLLDK 13

```

```

RESULT 20
US-08-300-928C-20
; Sequence 20, Application US/08300928C
; Patent No. 6019972
; GENERAL INFORMATION:
; APPLICANT: GEFTER, Malcolm L. et al.
; TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
; STREET: 610 LINCOLN STREET
; CITY: WALTHAM
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02145
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/300,928C
; FILING DATE: September 2, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,529
; FILING DATE: December 13, 1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: AMY E. MANDRAGOURAS
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: 002.6US(IMI-044)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
;
; US-08-300-928C-20

```

Query Match 2.4%, Score 6; DB 3; Length 19;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
 DB 8 LSLDK 13

RESULT 21
 US-08-430-944D-20
 ; Sequence 20; Application US/08430944D
 ; Patent No. 6025162
 ; GENERAL INFORMATION:
 ; APPLICANT: Bruce L. Rogers et al.
 ; TITLE OF INVENTION: A HUMAN T CELL REACTIVE FELINE PROTEIN
 ; NUMBER OF SEQUENCES: 103
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/430,944D
 FILING DATE: 28-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/430,014
 FILING DATE: 27-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/300,928
 FILING DATE: 02-SEPT-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Amy E. Mandragouras
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: IMI-044DV2
 TELEPHONE: (617)227-7400
 TELEFAX: (617)742-4214
 INFORMATION FOR SEQ ID NO: 20:
 LENGTH: 19 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal

US-08-430-944D-20
 Query Match 2.4%, Score 6; DB 3; Length 19;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
 DB 8 LSLDK 13

RESULT 22
 US-08-430-014-20
 ; Sequence 20; Application US/08430014
 ; Patent No. 6048562
 ; GENERAL INFORMATION:
 ; APPLICANT: GEFTER, Malcolm L. et al.
 ; TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE

; TITLE OF INVENTION: PROTEIN (TRFP)
 ; NUMBER OF SEQUENCES: 101
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
 ; STREET: 610 LINCOLN STREET
 ; CITY: WALTHAM
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02145

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/430,014
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/300,928
 FILING DATE:
 CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
 NAME: AMY E. MANDRAGOURAS
 REGISTRATION NUMBER: 36,207
 REFERENCE/DOCKET NUMBER: 002.6US(IMI-044)
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FRAGMENT TYPE: internal
 US-08-430-014-20

Query Match 2.4%, Score 6; DB 3; Length 19;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
 DB 8 LSLDK 13

RESULT 23
 US-08-431-184-20
 ; Sequence 20; Application US/08431184
 ; Patent No. 6120769
 ; GENERAL INFORMATION:
 ; APPLICANT: Bruce L. Rogers et al.
 ; TITLE OF INVENTION: A HUMAN T CELL REACTIVE FELINE PROTEIN
 ; NUMBER OF SEQUENCES: 103
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD, LLP
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/431,184
 FILING DATE: 28-APR-1995
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/430,014
FILING DATE: 27-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,928
FILING DATE: 02-SEPT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-044DV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-431-184-20

Query Match 2.4%; Score 6; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 LSLDDK 188
Db 8 LSLDDK 13

RESULT 24
PCT-US93-02462-10
Sequence 10, Application PC/TUS9302462
GENERAL INFORMATION:
APPLICANT: Gettel, Malcolm L.
APPLICANT: Garman, Richard D.
APPLICANT: Greenstein, Julia L.
APPLICANT: Kuo, Mei-chang
APPLICANT: Briner, Thomas J.
APPLICANT: Morville, Malcolm
TITLE OF INVENTION: PEPTIDES USEFUL FOR TOLERIZATION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/02462
FILING DATE: 19930325
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/006,116
FILING DATE: 15-JAN-1993
APPLICATION NUMBER: US 07/884,718
FILING DATE: 15-MAY-1992
APPLICATION NUMBER: 07/857,311
FILING DATE: 25-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-031PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
PCT-US93-02462-10

Query Match 2.4%; Score 6; DB 5; Length 19;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 LSLDDK 188
Db 8 LSLDDK 13

RESULT 25
US-07-807-529A-29
Sequence 29, Application US/07807529A
Patent No. 5547669
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce L.
APPLICANT: Morgenstern, Jay
APPLICANT: Bond, Julian F.
APPLICANT: Garman, Richard D.
APPLICANT: Greenstein, Julia L.
APPLICANT: Kuo, Mei-chang
APPLICANT: Morville, Malcolm
TITLE OF INVENTION: RECOMBITOPE PEPTIDES
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESSES:
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: One Kendall Square, Building 600
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02139
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807,529A
FILING DATE: 19911213
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,276
FILING DATE: 28-FEB-1991
APPLICATION NUMBER: US 07/431,565
FILING DATE: 03-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Channing, Stacey L.
REGISTRATION NUMBER: 31,095
REFERENCE/DOCKET NUMBER: IPC-027/Imi-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 494-0060
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-07-807-529A-29

Query Match 2.4%; Score 6; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
|||||
Db 1 LSLDK 6

RESULT 26
US-08-300-928C-78
; Sequence 78, Application US/08300928C
; Patent No. 6019972
; GENERAL INFORMATION:
; APPLICANT: GREYER, Malcolm L. et al.
; TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
; TITLE OF INVENTION: PROTEIN (TRFP)
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
; STREET: 610 LINCOLN STREET
; CITY: WALTHAM
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02145

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/300,928C
; FILING DATE: September 2, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,529
; FILING DATE: December 13, 1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: AMY E. MANDRAGOURAS
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: 002.605(IMI-044)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-300-928C-78

Query Match 2.4%; Score 6; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
|||||
Db 1 LSLDK 6

RESULT 27
US-08-430-944D-78
; Sequence 78, Application US/08430944D
; Patent No. 6023162
; GENERAL INFORMATION:
; APPLICANT: Bruce L. Rogers et al.
; TITLE OF INVENTION: A HUMAN T CELL REACTIVE FELINE PROTEIN
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston

; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,944D
; FILING DATE: 28-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/430,014
; FILING DATE: 27-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,928
; FILING DATE: 02-SEPT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-044DV2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-430-944D-78

Query Match 2.4%; Score 6; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLDK 188
|||||
Db 1 LSLDK 6

RESULT 28
US-08-430-014-78
; Sequence 78, Application US/08430014
; Patent No. 6048962
; GENERAL INFORMATION:
; APPLICANT: GREYER, Malcolm L. et al.
; TITLE OF INVENTION: PEPTIDES FOR HUMAN T CELL REACTIVE FELINE
; TITLE OF INVENTION: PROTEIN (TRFP)
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
; STREET: 610 LINCOLN STREET
; CITY: WALTHAM
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02145

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,014
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/300,928
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

NAME: AMY E. MANDRAGOURAS
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: 002.60US(IMI-044)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-430-014-78

Query Match 2.4%; Score 6; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 LSLLDK 188
Db 1 LSLLDK 6

RESULT 29
US-08-431-184-78
Sequence 78, Application US/08431184
Patent No. 6120769
GENERAL INFORMATION:
APPLICANT: Bruce L. Rogers et al.
TITLE OF INVENTION: A HUMAN T CELL REACTIVE FELINE PROTEIN
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,184
FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/430,014
FILING DATE: 27-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,928
FILING DATE: 02-SEPT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-044DV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)42-4214
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-431-184-78

Query Match 2.4%; Score 6; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 LSLLDK 188
Db 1 LSLLDK 6

RESULT 30
PCT-US93-02462-14
Sequence 14, Application PC/TUS9302462
GENERAL INFORMATION:
APPLICANT: Gifter, Malcolm L.
APPLICANT: Garman, Richard D.
APPLICANT: Greenstein, Julia L.
APPLICANT: Kuo, Mei-chang
APPLICANT: Briner, Thomas J.
APPLICANT: Morville, Malcolm
TITLE OF INVENTION: PEPTIDES USEFUL FOR TOLERIZATION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/02462
FILING DATE: 19930325
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/006,116
FILING DATE: 15-JAN-1993
APPLICATION NUMBER: US 07/884,718
FILING DATE: 15-MAY-1992
APPLICATION NUMBER: 07/857,311
FILING DATE: 25-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-031PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
PCT-US93-02462-14

Query Match 2.4%; Score 6; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 183 LSLLDK 188
Db 9 LSLLDK 14

RESULT 31
US-08-383-753-7
Sequence 7, Application US/08383753
Patent No. 5723584
GENERAL INFORMATION:

APPLICANT: Schatz, Peter J.
TITLE OF INVENTION: Biotinylation of Proteins
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: One Market Plaza, Steuart Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/383,753
FILING DATE: 03-FEB-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,991
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 1038.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-383-753-7

Query Match 2.4%; Score 6; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 QPVAVL 171
Db 2 QPVAVL 7

RESULT 32
US-08-586-772-7
Sequence 7, Application US/08586772
Patent No. 5874239
GENERAL INFORMATION:
APPLICANT: Schatz, Peter J.
TITLE OF INVENTION: Biotinylation of Proteins
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: One Market Plaza, Steuart Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/586,772
FILING DATE: 03-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/099,991
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 1038.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-586-772-7

Query Match 2.4%; Score 6; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 QPVAVL 171
Db 2 QPVAVL 7

RESULT 33
US-08-959-512-7
Sequence 7, Application US/08959512
Patent No. 5932433
GENERAL INFORMATION:
APPLICANT: Schatz, Peter J.
TITLE OF INVENTION: Biotinylation of Proteins
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: One Market Plaza, Steuart Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/959,512
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/383,753
FILING DATE: 03-FEB-1995
APPLICATION NUMBER: US 08/099,991
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 1038.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-959-512-7

Query Match 2.4%; Score 6; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 QPVAVL 171
|||||
Db 2 QPVAVL 7

RESULT 34
US-09-512-983-7
; Sequence 7, Application US/U9512983
; Patent No. 6265552
; GENERAL INFORMATION:
; APPLICANT: Scharz, Peter J.
; TITLE OF INVENTION: Biotinylation of Proteins
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/512.983
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/383.753
; FILING DATE: 03-FEB-1995
; APPLICATION NUMBER: US 08/099,991
; FILING DATE: 30-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 1038.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-512-983-7

Query Match 2.4%; Score 6; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 QPVAVL 171
|||||
Db 2 QPVAVL 7

RESULT 35
US-08-782-997A-24
; Sequence 24, Application US/08782997A
; Patent No. 6030602
; GENERAL INFORMATION:
; APPLICANT: Legendre, Jean-Yves
; ADDRESS: Supersaxo, Arnold
; FILING DATE: 28-FEB-1997

; TITLE OF INVENTION: Peptide Conjugates for Transfecting
; TITLE OF INVENTION: Cells
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782.997A
; FILING DATE: 14-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96100603.8
; FILING DATE: 17-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kreisler, Lewis J
; REGISTRATION NUMBER: 38,522
; REFERENCE/DOCKET NUMBER: RAN 4600/73
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (973) 235-4387
; TELEFAX: (973) 235-2363
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
US-08-782-997A-24

Query Match 2.4%; Score 6; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 VAVLGE 173
|||||
Db 26 VAVLGE 31

RESULT 36
US-08-807-332B-6
; Sequence 6, Application US/08807332B
; Patent No. 5959074
; GENERAL INFORMATION:
; APPLICANT: Dreyfus, David H.
; APPLICANT: Gelfand, Erwin W.
; TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
; TITLE OF INVENTION: GENE RECOMBINATION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,332B
; FILING DATE: 28-FEB-1997

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-807-332B-6

Query Match 2.4%; Score 6; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 GVLGTG 17
DB 32 GVLGTG 37

RESULT 37
US-09-338-876-6
Sequence 5, Application US/09338876
Patent No. 6187584
GENERAL INFORMATION:
APPLICANT: Dreyfus, David H.
APPLICANT: Gelfand, Erwin W.
TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,876
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,332
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-338-876-6

Query Match 2.4%; Score 6; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 GVLGTG 17
DB 32 GVLGTG 37

RESULT 38
US-07-662-193-3
Sequence 3, Application US/07662193
Patent No. 5328931
GENERAL INFORMATION:
APPLICANT: Kuo, Mei-chang
APPLICANT: Bond, Julian
TITLE OF INVENTION: Improved Preparation of Cat Dander
TITLE OF INVENTION: Allergens for Immunotherapeutic Purposes and Uses Thereof
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/662,193
FILING DATE: 19910228
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,276
FILING DATE: 28-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IML89-02AA/IPC-002CC/IMI-020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 70 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-07-662-193-3

Query Match 2.4%; Score 6; DB 1; Length 70;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLLDK 188
DB 58 LSLLDK 63

RESULT 39
5252466-3
Patent No. 5252466
APPLICANT: CRONAN, JOHN E.
TITLE OF INVENTION: FUSION PROTEINS HAVING A SITE FOR IN
VIVO POST-TRANSLATION MODIFICATION AND METHODS OF MAKING AND
PURIFYING THEM
NUMBER OF SEQUENCES: 24
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/525,568
FILING DATE: 18-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 354,266
FILING DATE: 19-MAY-1989
SEQ ID NO:3:
LENGTH: 75
5252466-3

Query Match 2.4%: Score 6; DB 6; Length 75;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 QPVAVL 171
Db 23 QPVAVL 28

RESULT 40

US-08-248-839C-67
Sequence 67, Application US/08248839C
Patent No. 5843702
GENERAL INFORMATION:
APPLICANT: McGonnell, David
APPLICANT: Devine, Kevin
APPLICANT: O'Kane, Charles
TITLE OF INVENTION: A Gene Expression System
NUMBER OF SEQUENCES: 185
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5843702o No. 5843702disk of No. 5843702th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,839C
FILING DATE: 25-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3614,214-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-248-839C-67

Query Match 2.4%: Score 6; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 LAQOKA 78
Db 50 LAQOKA 55

RESULT 41

5459046-11

Patent No. 5459046
APPLICANT: KODAMA, TOHRU-IGARASHI, YASUO
TITLE OF INVENTION: CYTOCHROME C GENE DERIVED FROM HYDROGEN
BACTERIUM
NUMBER OF SEQUENCES: 17
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/943,140
FILING DATE: 10-SEP-1992
PRIOR APPLICATION DATA:486,409
APPLICATION NUMBER: 28-FEB-1990
SEQ ID NO:11:
LENGTH: 80
5459046-11

Query Match 2.4%: Score 6; DB 6; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 QLAKQK 77
Db 3 QLAKQK 8

RESULT 42

US-08-855-5310-9
Sequence 9, Application US/08855531D
Patent No. 6110467
GENERAL INFORMATION:
APPLICANT: PREM, PAUL S.
HALBUR, PATRICK G.
MENG, XIANG-JIN
LUM, MELISSA A.
LYOO, YOUNG S.
TITLE OF INVENTION: VACCINES RAISING AN IMMUNOLOGICAL
RESPONSE AGAINST VIRUSES CAUSING PORCINE RESPIRATORY AND
REPRODUCTIVE DISEASES, METHODS OF PROTECTING A PIG AGAIN
A DISEASE CAUSED BY A RESPIRATORY AND REPRODUCTIVE

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/855,531D

FILING DATE: 13-May-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/969,071

FILING DATE: 30-OCT-1992

ATTORNEY/AGENT INFORMATION:

NAME: LAVALLEY, JEAN-PAUL M.P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 4625-038-55X DIV

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 85 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-08-855-531D-9

Query Match 2.4%; Score 6; DB 3; Length 85;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 VTANVT 134
Db 41 VTANVT 46

RESULT 43

US-08-855-526B-9
; Sequence 9: Application US/08855526B
; Patent No. 6251404
; GENERAL INFORMATION:
; APPLICANT: PREM, PAUL S.
; APPLICANT: HALBUR, PATRICK G.
; APPLICANT: MENG, XIANG-JIN
; APPLICANT: LUM, MELISSA A.
; APPLICANT: LYOO, YOUNG S.
; TITLE OF INVENTION: VACCINES RAISING AN IMMUNOLOGICAL
; TITLE OF INVENTION: RESPONSE AGAINST VIRUSES CAUSING PORCINE RESPIRATORY AND
; TITLE OF INVENTION: REPRODUCTIVE DISEASES, METHODS OF PROTECTING A PIG AGAINST
; TITLE OF INVENTION: A DISEASE CAUSED BY A RESPIRATORY AND REPRODUCTIVE
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ORLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,526B
; FILING DATE: 13-MAY-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/969,071
; FILING DATE: 30-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: LAVALLEYE, JEAN-PAUL M.P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 4625-040-55X DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-855-526B-9

Query Match 2.4%; Score 6; DB 4; Length 85;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 VTANVT 134
Db 41 VTANVT 46

RESULT 44

US-07-807-529A-4
; Sequence 4: Application US/07807529A
; Patent No. 5547669
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce L.
; APPLICANT: Morgenstern, Jay
; APPLICANT: Bond, Julian F.
; APPLICANT: Garman, Richard D.
; APPLICANT: Greenstein, Julia L.
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Morville, Malcolm
; TITLE OF INVENTION: RECOMBITOPE PEPTIDES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
; STREET: One Kendall Square, Building 600
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02139
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,529A
; FILING DATE: 19911213
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,276
; FILING DATE: 28-FEB-1991
; APPLICATION NUMBER: US 07/431,565
; FILING DATE: 03-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Channing, Stacey L.
; REGISTRATION NUMBER: 31,095
; REFERENCE/DOCKET NUMBER: IPC-027/imi-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 494-0060
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-807-529A-4

Query Match 2.4%; Score 6; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 183 LSLLDK 188
Db 76 LSLLDK 81

RESULT 45

PCT-US93-02462-4
; Sequence 4: Application PC/TUS9302462
; GENERAL INFORMATION:
; APPLICANT: Gefter, Malcolm L.
; APPLICANT: Garman, Richard D.
; APPLICANT: Greenstein, Julia L.
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Briner, Thomas J.
; APPLICANT: Morville, Malcolm
; TITLE OF INVENTION: PEPTIDES USEFUL FOR TOLERIZATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02462
; FILING DATE: 19930325
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/006,116
; FILING DATE: 15-JAN-1993
; APPLICATION NUMBER: US 07/884,718
; FILING DATE: 15-MAY-1992
; APPLICATION NUMBER: 07/857,311
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-031PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-02462-4

```

```

Query Match          2.4%; Score 6; DB 5; Length 88;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 183 LSLLOK 188
    |||||
DB 76 LSLLOK 81

```

```

RESULT 46
US-08-997-080-143
; Sequence 143, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-997-080-143

```

```

Query Match          2.4%; Score 6; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 129 VTANVT 134
    |||||
DB 45 VTANVT 50

```

```

RESULT 47
US-08-997-362-143
; Sequence 143, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul.
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 amino acids

```

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-362-143

Query Match 2.4%; Score 6; DB 2; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 VTANVT 134
DB 45 VTANVT 50

RESULT 48

US-09-095-855-143
Sequence 143, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:

ADDRESSEE: law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:

INFORMATION FOR SEQ ID NO: 143:

SEQUENCE CHARACTERISTICS:
LENGTH: 91 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-095-855-143

Query Match 2.4%; Score 6; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 VTANVT 134
DB 45 VTANVT 50

DB 45 VTANVT 50

RESULT 49

US-09-324-542-143
Sequence 143, Application US/09324542
Patent No. 6328978
GENERAL INFORMATION:

APPLICANT: Watson, James D.
APPLICANT: Tan, Paul L.J.
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Methods and Compounds for the Treatment
FILE REFERENCE: 11000.1007c1
CURRENT APPLICATION NUMBER: US/09/324,542
CURRENT FILING DATE: 1999-06-02
EARLIER APPLICATION NUMBER: US 08/997,080
EARLIER FILING DATE: 1997-12-23
NUMBER OF SEQ ID NOS: 194
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 143
LENGTH: 91
TYPE: PRT
ORGANISM: Mycobacterium vaccae
US-09-324-542-143

Query Match

2.4%; Score 6; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 VTANVT 134
DB 45 VTANVT 50

RESULT 50

US-07-807-529A-2
Sequence 2, Application US/07807529A
Patent No. 5547669
GENERAL INFORMATION:

APPLICANT: Rogers, Bruce L.
APPLICANT: Morgenstern, Jay
APPLICANT: Bond, Julian F.
APPLICANT: Garman, Richard D.
APPLICANT: Greenstein, Julia L.
APPLICANT: Kuo, Mei-chang
APPLICANT: Morville, Malcolm
TITLE OF INVENTION: RECOMBITOPE PEPTIDES
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
STREET: One Kendall Square, Building 600
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02139

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807,529A
FILING DATE: 19911213
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/662,276
FILING DATE: 28-FEB-1991
APPLICATION NUMBER: US 07/431,565
FILING DATE: 03-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Channing, Stacey L.

REGISTRATION NUMBER: 31,095
REFERENCE/DOCKET NUMBER: IPC-027/im1-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 494-0060
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 92 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-07-807-529A-2

Query Match 2.4%; Score 6; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 LSLIDK 188
IIIIII
Db 80 LSLIDK 85

Search completed: August 6, 2002, 16:46:09
Job time: 371 sec